

Don't blame NASA alone for Mars mission failures

While the spotlight has been focused on mistakes by US space engineers and managers for the loss of recent Mars-bound spacecraft, others share responsibility. This includes those who have set unrealistic goals for the agency.

Eight years into his tenure as the longest-serving administrator for the US space agency NASA, Dan Goldin has hit a bad patch of road. Last year's embarrassing Mars failures have subjected his cherished 'better, faster, cheaper' philosophy to second-guessing, and the administrator and his deputies have responded somewhat defensively by tallying up the agency's win-loss record since 1992: 146 payloads launched, only 10 lost. See? The strategy works.

They should relax. No one, least of all the panel led by aerospace executive Thomas Young that delivered last week's critique of the Mars programme, is calling for a return to big, infrequent missions (see page 535). And most people are willing to accept some additional risk to boost the flight rate. Most would also agree that Goldin is a visionary, and that his transformation of NASA has been truly revolutionary. But he is also a politician. And NASA tends to run into trouble when politics is injected into a business that is difficult even under the best circumstances.

It would be wrong to say the Mars programme was politically motivated. The planet's similarity to Earth and the existence of water (certainly in the past, maybe under the surface today) make it a compelling scientific destination. But the 1996 claim of fossils in a Martian rock added a public relations (hence political) dimension to the programme that hadn't been there before. This, along with Goldin's eagerness to prove that his employees could do more with less, are what led to the downfall.

The Young report treats this NASA pressure on the Jet Propulsion Laboratory (JPL) delicately: "NASA Headquarters thought it was

articulating programme objectives, mission requirements, and constraints. JPL management was hearing these as non-negotiable programme mandates." This may be letting NASA off the hook too easily. The political realities of the Mars programme were not misinterpreted by the JPL engineers and scientists in the trenches. They understood that they could not ask for more money, nor could they radically 'descope' their missions. Their only choice was to sigh and accept more risk. That, or resign.

Goldin showed class last week by taking the dejected JPL Mars team to dinner, then repeating his *mea culpa* in public: "I pushed too hard, and in doing so stretched the system too thin."

Others have yet to admit their share of the blame. The White House, which scrutinizes and approves every NASA budget request, said not a word last week. James Sensenbrenner, who chairs the agency's authorizing committee in the House of Representatives, issued his own defensive statement pointing out that Congress has given NASA more money than it requested for space science in five of the past six years.

Do Goldin and those who determine his funding really understand where they went wrong? In the case of the Mars programme, probably yes. But it will be interesting to see how they respond to another current agency project that has the same disturbing combination of engineering complexity, too few resources, politics that constrain project managers' decisions, and workers just trying to make the best of a bad situation. It's called the International Space Station. ■

Taking the initiative

Two new sections of *Nature* should help to keep readers informed and up to date on important areas of science.

Publications whose staff sit back and wait for the world to come knocking on their door are likely to be short-lived. Nevertheless, *Nature* could easily fill its gradually increasing number of pages every week by being reactive rather than proactive. Events would happen and journalists would write about them; we would go on publishing the best of what we receive in the form of original papers, and News and Views authors would explain them.

But that philosophy would do less than justice to our readers, who deserve more by way of editorial initiative. *Nature* has never lacked proactivity. But this week sees two significant further steps in that spirit: the launches of the News Features and *Nature* Insight sections (see pages 538 and 631, respectively). *Nature* already aims to publish overviews that are authoritative, timely, dynamic and, above all, accessible. Our News Features represent an enhancement of our regular journalistic coverage that is intended to help meet this goal.

The *Nature* Insight section, to be published every month, will bring together a collection of articles covering many aspects of a

timely subject, in the form of commentaries and review articles written by acclaimed scientists in the field. As the boundaries between disciplines are blurring — for example, with physicists trying to understand intricate details of cells and biologists modelling global signalling networks — there has never been a greater need to understand specialist issues across the expanse of biology and the physical sciences. Claiming to provide a one-stop shop on any subject would smack of hubris, but that is what we aspire to. The reviews are presented in a style designed for maximum clarity, with special attention to annotated schematics to aid understanding of the text, and boxes containing information for the specialist or background information for the general reader. Thus, in contrast to traditional review articles, we hope that these collections of overviews will entice readers not already closely involved in the area of research.

In this issue we have focused both the *Nature* Insight and the News Feature on the same topic. We hope readers will enjoy the complementarity of these approaches. ■





US biomedicine
Congress should
address controversy,
says Harold Varmus
p530



Spyhole
Research aircraft
detect deterioration
of Arctic ozone hole
p531



Whale song
Fears for grey whale
population under
Japanese proposals
p532



Mars mission
Plans to explore the
Red Planet are being
cut back
p535

Complaints grow over delays in UK animal licence processing

London

The Research Defence Society (RDS), which represents the interests of British scientists who use animals in their research, used to get a trickle of complaints about delays in granting licences for experiments using animals. Now, it claims, that trickle has become a flood of around 30 complaints a month.

Britain operates arguably the world's toughest regulations governing research on animals. Many scientists support the goal of improving animal welfare. But they claim that the administrative burden of processing licence applications, amendments and renewals has increased beyond the system's capacity to cope.

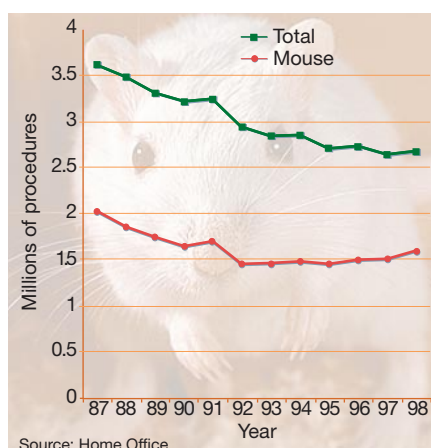
The situation has worsened since April 1999, when further regulations were introduced and a new layer was added to the process: local ethical review boards. Delays in processing applications in the Home Office — the government department that approves licences for animal research — were already common.

But they seem to be getting worse, and the new hurdles have stretched scientists' patience. "I am on the verge of giving up animal research because of the bureaucratic cost," says one physiologist.

While feelings are running high, few scientists are prepared to speak out openly. Years of violent protests from extremists at the fringe of the British animal-rights movement have left a climate of fear.

Colin Page, a medical researcher at King's College London, has already fallen victim to intimidation. In 1998, when an animal-rights activist jailed for arson was threatening to starve himself to death, a group called the Animal Rights Militia promised to kill ten named scientists if he died. Page was number two on the list.

Page works on lung disease, and is trying to find out how white blood cells get into lung tissue. One approach is to use genetic 'knock-out' mice that lack the proteins involved, but the time taken to gain a licence to do the work in Britain means Page cannot compete with researchers elsewhere. "We are talking about sending someone to the United States as a faster way of doing this work," he says.



On alert: the rising popularity of transgenic mice is leading to an increase in animal experiments.

Similar complaints are not hard to find. One researcher applying to renew a project licence waited "an astonishing and depressing 15 months". In another example, a veterinary practitioner who started a PhD at a university

in northeast England has not been paid since February because funds from his studentship will not arrive until his licence is approved.

The Home Office is questioning the number of animals he plans to use. "I don't want to see animals wasted," says the researcher. "But there is no point doing three years' work and finding you haven't enough animals to get statistical significance."

Les Ward, director of Edinburgh-based Advocates for Animals, which campaigns for the abolition of animal experiments, sees things differently. "Those complaining of bureaucracy need to appreciate that it's for the best reasons: animal welfare, and making sure the best science is used to reassure the public in whose name it is being done."

Ward, who sits on the Home Office's Animal Procedures Committee, accuses the RDS of "spinning" the issue by encouraging researchers to complain about such administrative delays. "We have some die-hard scientists who resent any change," he adds. "These people have to accept that their applications

Japan sets tissue donor guidelines

Tokyo

Japan's Ministry of Health and Welfare last week released a set of guidelines covering the transplantation and medical use of human cells and tissues. The guidelines specify safety-assessment procedures and informed-consent rules for donated cells and tissues that are not regulated by Japan's organ transplant law.

In contrast to the transplant law, which requires written consent by the donor, the new guidelines specify that cells or tissues can be taken from dead donors if informed consent has been obtained from family members.

Scientists have welcomed the guidelines as confirming that surplus tissue — as well as tissue that has proved unsuitable for medical use — may be used for research as long as informed consent has been obtained. This was not allowed under the organ

transplant law, which specifies that all surplus tissue must be burned.

But Toshiharu Matsumura, a general manager at the Meiji Cell Technology Center, argues that, in the long term, a clear separation of medical and research use may be inevitable. He says it would be more appropriate to regulate the provision of human materials by law rather than through guidelines.

"Without a legal basis, it may turn out rather difficult to build up an efficient support system for human tissues, which are increasingly crucial in many fields of biomedical research," he says. Some observers argue that public sensitivity towards collecting tissues from dead bodies, along with media criticism, could lead to a backlash, resulting in strict rules — as was the case with organ transplants.

Robert Triendl

will not be rubber-stamped any more.”

Many scientists, however, accept that some proposals will be turned down on ethical grounds. But they want to know quickly, so they can move on to other projects. The main reason for the delays, they argue, is that the number of officials has failed to keep pace with the burgeoning paperwork.

Some universities are hiring staff to cope with the local review. But the Home Office has only 21 inspectors to oversee the entire regulatory system. “You must cater for the changes by increasing the number of inspectors,” says the veterinary researcher.

Some animal-welfare groups agree. The Fund for the Replacement of Animals in Medical Experiments, based in Nottingham, says inspectors turn round applications as quickly as they can — but their workloads are simply too high, and are likely to increase.

Over the past 25 years, the number of experimental procedures involving animals in Britain has decreased steadily. But last year, the total rose by 1 per cent, largely due to a growth in demand for transgenic mice.

As more researchers turn to transgenic mice, this trend is likely to increase. RDS director Mark Matfield predicts that the rising demand for transgenics could push up the total number of animal procedures in Britain by 5 per cent a year. This will be embarrassing for the government, which is committed to reducing the number of animals used for research.

“We are planning to review the ethical review procedure,” says a Home Office spokesman. “We are aware of concerns expressed, but we remain convinced that it has a positive role to play.”

Natasha Loder

Varmus tells Congress to grasp thorny policy issues

Washington

Former director of the National Institutes of Health, Harold Varmus, returned to Washington briefly last week to tell a group of congressmen that they should spend more time addressing the broad political issues that affect biomedical research — and less time seeking budget increases for popular medical causes.

High on the agenda of those interested in promoting such research, he suggested, should be the ‘commodification’ of databases — a reference to the activities of companies that generate and repackage scientific data and sell it to subscribers.

Varmus was addressing the tenth anniversary celebrations of the US Congress-

sional Biomedical Caucus. Partly as a result of the success of the caucus, the NIH has received 15 per cent budget increases in each of the past two years, and congressional leaders are pushing for a third such rise.

But Varmus, who left the NIH to become president of the Memorial Sloan-Kettering Cancer Research Center this year, is worried that there may be an imbalance in priorities. “We should take the issues on when they are hot,” he said. Issues that the caucus should address include the sale of human genetic information, lapses in overseeing clinical research, and the growth of online research literature, he said.

In its recent activities, the caucus has preferred to focus on less contentious problems, such as the need to combat cystic fibrosis. The failure to discuss more controversial issues leads to misunderstandings when they inevitably arise, Varmus said. He predicted, for example, that the availability of data will become an increasing problem as genomics becomes a bigger business. “Sets of data have become commodities,” he said. “This is going to affect the relationship between public and private sectors.”

But despite Varmus’s urgings, members of the caucus seemed more comfortable repeating conventional concerns, and confirming their belief that the main role of the caucus is to generate enthusiasm for science funding.

Paul Smaglik



Varmus: wants caucus to reassess priorities.

US geneticists encouraged to play by the book

Washington

US geneticists were warned last week by their professional body that they could be endangering their research if they ignore government regulations protecting the families of subjects involved in human genetics studies.

But the American Society of Human Genetics (ASHG) also suggested that researchers should learn to make use of waivers, or face the “enormously cumbersome and prohibitive” task of enrolling family members as human subjects, a burden that would “seriously impede medical research”.

In an alert sent out to its 6,700 members, the society drew attention to a recent case that led to the closure of federally funded research at Virginia Commonwealth University (VCU) in Richmond. The father of college-age twins had complained that a questionnaire sent to his daughter by a

university geneticist invaded his privacy by asking hundreds of questions about his own physical and emotional health (see *Nature* 404, 114–115; 2000).

The Office for Protection from Research Risks (OPRR) of the National Institutes of Health ruled that the university’s ethics board should have required the researcher to obtain informed consent not only from recipients of the questionnaire, but also from members of their families.

The alternative, said the OPRR, was that the ethics board should have formally waived this requirement by agreeing that the research posed minimal risk to the rights of other family members and could not be practicably carried out without the waiver.

The ASHG alert was posted by the society’s immediate past president, Uta Francke, a professor of genetics at Stanford University. She says that before the VCU case she “didn’t know a thing” about the

federal regulations covering the collection of information about family members of human subjects, and assumed that many geneticists were equally ignorant. “To prevent any [incidents similar to the VCU case] in the future, our members need to be aware that these regulations exist,” she says.

The society’s approach contrasts with the more critical stance of Francis Collins, the director of the National Human Genome Research Institute, who challenged the OPRR’s application of federal rules at VCU, declaring his “deep concern” about its possible implications for genetics research.

The ASHG alert, in contrast, appears to embrace the ruling, declaring that “there are clear lessons to be learned” from it. In particular, it stresses that geneticists should know the conditions required for ethics boards to grant waivers of informed consent from family members.

Meredith Wadman



Into the vortex: NASA 'spy' planes have detected renewed ozone depletion in the Arctic stratosphere.

Japan squares up to conservationists over grey whale's status

Tokyo

The endangered Asian population of the grey whale could be threatened by Japanese proposals to downgrade the level of protection offered to the larger northeastern Pacific population, wildlife conservationists and researchers have warned.

Concern has been heightened by an analysis of whale products in Japanese shops, which showed that some of the meat on sale came from grey whales. The legal commercial hunting and trading of whales ended in 1986, although Japan is still allowed to hunt whales for research purposes.

Japan's Fisheries Agency wants to downlist three whale populations — the South Antarctic and northwestern Pacific Minke whale and northeastern Pacific grey whale. It plans to table the proposals at the eleventh meeting of the Convention for International Trade in Endangered Species of Wild Fauna and Flora, which starts next week in Nairobi, Kenya.

The agency says the proposals are based on research by the International Whaling Commission, which showed that the populations of these species are abundant. For example, the American grey whale population now totals more than 22,000. But, conservationists point out, the Asian population is estimated at no more than 100 individuals.

According to researchers from the University of Auckland, DNA analysis of the products on sale in Wakayama prefecture, on Japan's Pacific coast, showed that they came from a grey whale. But the researchers could not specify which population the DNA came from — northeastern Pacific, American or Asian.

"The haplotype match could be explained by the low genetic diversity among grey whale populations," says Scott Baker, leader of the University of Auckland team. He explains that variation among Asian grey whales is low because of the small population size, whereas the 'bottleneck' — a period of reduced population size — caused by hunting of the American grey whale is thought to account for the lack of variation in the current population. "But the absence of a comparative genetic study between the two populations makes it difficult to draw any conclusion at this stage."

Past requests from researchers for tissue samples or genetic data from stranded whales killed by accident have been turned down repeatedly by Japanese

Global warming could be bad news for Arctic ozone layer

London

Environmental scientists have struggled for years to dispel a popular misconception that global warming and the thinning ozone layer are one and the same. But researchers who have just completed the largest ever survey of Arctic stratospheric ozone now warn that the two processes are, indeed, connected.

The bad news is that climate change could prolong ozone depletion in the Arctic by many years — despite the success of the Montreal Protocol of 1987 in reducing emissions of ozone-destroying chemicals.

Preliminary results, released this week, indicate that Arctic ozone loss is back at the alarming levels seen in the mid-1990s, when three successive winters, from 1994–95 to 1996–97, raised serious concerns. This year, at altitudes of around 18 kilometres, more than 60 per cent of the ozone was destroyed.

If the polar vortex — a stable mass of cold air that forms above the Arctic each winter — gets cold enough, icy clouds form in the arid stratosphere. These polar stratospheric clouds harbour compounds containing chlorine and bromine, which destroy ozone.

For the past few months, that process has been under unprecedented scrutiny as more than 500 scientists have used a suite of research aircraft, balloons, satellites and ground-based instruments to monitor ozone depletion in the polar vortex.

The survey, which has just finished, was a collaboration between the European Commission, which funded the THESEO 2000 programme, and the United States, which called its project SOLVE. As *Nature* went to press, THESEO 2000 scientists were attempting a final balloon flight.

As part of the survey, an ER-2 aircraft — a civilian version of the U2 spy plane, from the US space agency NASA — flew from Kiruna

in Sweden into the vortex, recording the concentration of ozone-destroying compounds and a wide spectrum of chemicals including nitrous oxide, carbon dioxide and methane.

Using these measurements, scientists were able to 'fingerprint' discrete parcels of stratospheric air, and then sample them repeatedly over a period of time with a range of instruments. These analyses revealed steady declines in ozone levels from January.

The extent of ozone destruction in a given year depends on how cold the vortex gets, and how long it lasts. In both 1997–98 and 1998–99, the vortex was a few degrees warmer than in the middle of the decade, and broke up before the spring, when increasing daylight hastens the chemical reactions that destroy ozone. As result, the Arctic retained most of its ozone.

The damage can also be limited by the presence in the stratosphere of nitric acid, which combines with chlorine and prevents it attacking ozone. But if ice particles in the polar stratospheric clouds sink down from the stratosphere, they take nitric acid with them. That happened this year. "There was extensive denitrification in the polar vortex," says Paul Newman of NASA's Goddard Space Flight Center in Greenbelt, Maryland.

The worry is that global warming will make this year's pattern more common. Most modellers believe that the Arctic stratosphere will cool as the lower atmosphere warms. Certainly, winters like this year's, in which the polar vortex has been cold and long-lasting, seem to be becoming more frequent. "One year does not prove a case," says Newman. "But we have seen quite a few years lately in which the stratosphere has been colder than normal."

Peter Aldhous

♦ <http://www.ozone-sec.ch.cam.ac.uk>

♦ <http://cloud1.arc.nasa.gov/solve/>



Unkind cuts: despite a hunting ban, grey whale meat may still be ending up in shops.

authorities. Some link this to Japanese regulations that allow the local use of meat from whales stranded or entangled in fishing gear regardless of their protective status.

The team believes it is possible that the products came from an Asian grey whale found floating in the Sea of Japan near the northern island of Hokkaido in 1996. Although the whale was reported as being 'stranded', a later investigation revealed that it had been harpooned, with a 5–6 metre section of its tail expertly severed.

"It is possible that the products originate from long-term stockpiling of whales killed before 1986," says Baker. "But if they prove to be derived from the Hokkaido whale, the practice of stockpiling applies to illegal products, and the meat is being sent to places which could hardly be described as 'local'.

"Either way, undocumented and possibly illegal products from endangered species are turning up in the commercial market, and allowing trade in products from the American population should not be considered until this lack of transparency is solved."

The Fisheries Agency insists its proposed downlisting of the three whale species is not intended to lead to a resumption of trade in whale meat, but rather to pursue appropriate management of the species based on scientific findings.

But Naoko Nakamae, the Japanese representative of the International Fund for Animal Welfare, says such a move would put the Asian stock of grey whales at risk. "Unless the Japanese authority decides to make its data available to all, differentiating the two populations will remain impossible," she says. **Asako Saegusa**

US court tests the breadth of patent protection on proteins

Washington

Does a patent on a recombinantly produced protein give the owner rights to versions of that protein produced in other ways? This issue is being put to the test in a closely watched battle that opened in a Boston courtroom last week.

The biotechnology company Amgen, based in Thousand Oaks, California, is suing Transkaryotic Therapies (TKT), a small company in Cambridge, Massachusetts, alleging infringement of several patents on its blockbuster drug Epogen. Used to treat anaemia, Epogen is the trade name for erythropoietin (EPO), the hormone that stimulates the production of red blood cells.

Amgen developed the drug by cloning the human EPO gene and inserting it in Chinese hamster ovary cells, which then produce the hormone. US sales of Epogen for treating anaemia in kidney-dialysis patients totalled \$1.8 billion last year. Johnson & Johnson, which sells Epogen for other uses under licence to Amgen, generated an additional \$2 billion in sales. Worldwide, the Epogen market is estimated to be worth \$5 billion a year.

Enter TKT. The company's researchers have developed an alternative way to produce EPO, taking advantage of the fact that all human cells carry the EPO gene, even though it is active only in kidney cells. In TKT's process, a promoter — or 'on-switch' — for the gene is inserted in human cells along with other regulatory and structural elements that guide the promoter to the EPO gene. The gene becomes activated and the cells begin producing the hormone.

TKT plans to market EPO produced in this way in collaboration with Aventis, the giant European drug-maker. It claims that its production process is distinct from Amgen's, so it is not infringing on the company's patents on the EPO gene and EPO production. Indeed, TKT's process does not require knowledge of the gene sequence.

More importantly, the Cambridge company disputes Amgen's patent rights to the EPO protein itself when made by mammalian cells in culture — rights that will be at the centre of the battle in US District Court. TKT and its advocates argue that Amgen is asserting far too broad a claim to a naturally occurring protein that was well characterized decades before Amgen cloned the gene and was granted its first patent in 1987. Indeed, Amgen scientists worked backwards from the protein's known amino-acid sequence to deduce the gene sequence.

If TKT prevails against Amgen, the implications would be considerable for Amgen's lucrative EPO market. They could also be felt

by companies that have commercialized other therapeutic proteins such as interferon and Neupogen, another Amgen product.

TKT says that it has six additional proteins in its therapeutic pipeline, all made by the same promoter-insertion process. If it were to bring any of them to market — and presuming it prevailed in the inevitable court battles that would ensue — it could make considerable dents in what until now have been relatively exclusive and lucrative markets.

As a result, if TKT wins the lawsuit, "the industry may kind of go into shock," says Eric Schmidt, a biotechnology analyst at S. G. Cowen Securities in New York.

Amgen itself predicts dire consequences in such an event. "The real question is how strong is the US patent system," Gordon Binder, Amgen's chief executive, told *The New York Times* after the first day of preliminary court proceedings last week.

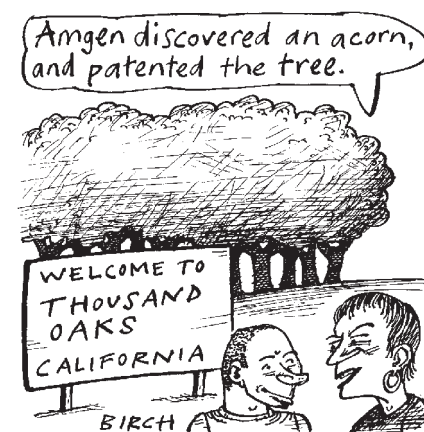
Some analysts back Amgen. "We believe that the Amgen patents do cover the TKT EPO product, if not the production process," says Dennis Harp of Deutsche Banc Alex. Brown, a global financial services company.

But others disagree. "If Amgen prevails, it suggests that US patents are so restrictive that it would actually discourage legitimate innovation," says Jonas Alsenas, an analyst at IND Bearings in New York. "If by solving half the problem you get rights to the full solution to the problem, that's not fair."

TKT declines to comment because the case is in litigation. Earlier this year, Richard Selden, its chief executive, told *The New York Times*: "We think that we're doing something fundamentally different" from Amgen.

David Kaye, an Amgen spokesman, says "We're confident in our patent position". The full trial of the case, which Amgen filed in 1997, is expected to begin in May. Whatever verdict is reached is almost certain to be appealed against.

Meredith Wadman



Baja peninsula claims five victims in tragic accident

San Diego

The ecological research community is in shock following the death of five of its members. The researchers were on an expedition in the Gulf of California, off Mexico's Baja peninsula, when their boat was caught in a sudden wind storm.

Among the dead are Gary Polis, chairman of the department of environmental science and policy at the University of California at Davis, and three Japanese researchers from Kyoto University: Takuya Abe, Masahiko Higashi and Shigeru Nakano, who is presumed dead although his body had not been recovered by last weekend. Michael Reese, a postgraduate ecologist at Davis, also drowned.

Four other members of the team — three graduate students and a postgraduate — were rescued after swimming for nearly four hours to islands.

Polis, a zoologist known for his research on scorpions, was also president of the American Society of Naturalists. Abe and Nakano were known for their work on biodiversity in the tropical forests of



Polis: an experienced ecologist who leaves a "bright legacy of accomplishments".

Borneo; Higashi was a specialist in food-web theory.

"The researchers who died left us a bright legacy of accomplishments that will make them remembered," says Richard Atkinson, president of the University of California.

The Baja peninsula, which stretches for 950 sparsely populated kilometres down Mexico's northwest coast, is home to a wide range of plants and animals. But its desolate

beauty masks many dangers, such as the sudden winds that can sweep off Baja's mountainous desert at up to 80 km per hour, whipping the water into huge waves.

The research team was operating about 500 km south of San Diego, near the remote fishing village of Bahia del Los Angeles on Baja's eastern shore, an area Polis knew well. They left the village in the morning in two motorized boats to do fieldwork on the Isla de Cabeza de Caballo, about 6 km offshore.

The open boats were returning to the village at midday when disaster struck. The boats became separated in the high seas, and one was swamped.

The researchers' project — examining spiders and scorpions — is funded by the US National Science Foundation and the Earthwatch Institute.

The Japanese researchers were building links between Kyoto and Davis, both specifically related to the project and more generally. They arrived in Davis the previous week and had spent a few days at the university before driving south into Mexico.

Rex Dalton & David Cyranoski

Reshuffle lifts French synchrotron hopes

Paris

The sacking of France's research minister, Claude Allègre, a fortnight ago has cast uncertainty over the level of France's promised participation in the construction of the British synchrotron Diamond — and has reawakened hopes that a French machine might also be built.

Government officials in Paris say that France's involvement in the £200 million (US\$313 million) Diamond, which is being jointly financed by the British government and the Wellcome Trust, has not been sealed by any formal agreement. The controversial decision to participate was taken by Allègre last year, ending eight years of plans to build a French machine, Soleil (see *Nature* 400, 489; 1999).

France's new science minister, Roger-Gérard Schwarzenberg, has not made any official statement about Soleil. But he has promised to study the dossier, together with a recent parliamentary report urging the science ministry to push ahead with the construction of Soleil "without further delay" (see *Nature* 404, 323; 2000).

Allègre had argued that Soleil would be too expensive to build and that large scientific equipment should be built at the European level. But his decision, which was strongly criticized by France's scientific and political communities, was not sealed by any formal

agreement between the two governments and the Wellcome Trust, according to Vincent Courtillot, research director under Allègre.

In principle, Schwarzenberg could reduce France's financial investment in Diamond so as to proceed with the French project — France is unlikely to be able to afford the FF350 million (US\$51 million) announced by Allègre for Diamond as well as pay for a new synchrotron on its own soil.

Another option would be for France to find other partners for Soleil. Belgium, Holland and Spain are possible candidates, says Courtillot, who is expected to remain on Schwarzenberg's team.

Senator René Trégouët (RPR, Rhône), an author of the parliamentary report, said last week that France should remain flexible over Diamond so as to have access to some beamlines on the machine, while also being able to move ahead with a French synchrotron.

Trégouët says French researchers will need at least ten beamlines over the next decade, and that the most economic solution would be to rent beamlines on Diamond and build a national machine with a foreign partner.

"The decision [to build] Soleil should be taken very rapidly. We need to construct a machine in France and we need to have an accord, perhaps between England, Germany and France," says Trégouët.



Schwarzenberg: a reprieve for Soleil?

Schwarzenberg, a professor of law at the University of Paris and a seasoned politician, is relatively unknown to the scientific community, and would gain an immediate political boost by resurrecting Soleil.

"We hope with this new minister that we will have [our own synchrotron] source," says Bernard Capelle, director of research at the Centre National de la Recherche Scientifique and a member of the tripartite group planning Diamond with the Wellcome Trust and the British government.

"It probably will not be a completely French source, and we might have to reduce our participation in Diamond. But the two machines would be complementary, with the French machine having a lower energy," Capelle says.

A spokesperson for Britain's Central Laboratory for the Research Councils, which will run Diamond, declined to comment. The issue was believed due to be discussed with French officials this week.

Heather McCabe

Cereal gene bank accepts need for patents...

Obregon, Mexico

The world's leading maize and wheat research organization last week adopted an intellectual-property policy intended to ensure that its resources will remain available to scientists working for sustainable agriculture in developing nations.

The international board of trustees of the Mexico-based International Maize & Wheat Improvement Centre (CIMMYT) — which holds the world's largest maize and wheat gene bank — unanimously approved the policy to prevent private companies from claiming intellectual-property rights over any of its discoveries or resources.

The centre fears its long-standing mission of providing germplasm free to plant breeders worldwide would be threatened without the policy. CIMMYT previously did not have a policy of protecting its research discoveries.

"Our desire is to ensure that inventions aren't patented out from under the organization," says Wallace Falcon, a Stanford University economist and CIMMYT board chairman.

CIMMYT board member Cary Fowler, a sociologist at the Agricultural Institute of Norway, cautioned that this is not an effort by the organization to "get rich" by patenting discoveries, but to ensure broad distribution of plant materials through a flexible policy.

Similar steps are being considered by other members of the Consultative Group on International Agricultural Research (CGIAR), an umbrella organization that supports sustainable agriculture for developing nations.

Last week, the Philippines-based International Rice Research Institute (IRRI) organized a conference in Los Baños of the 12-nation Council for Partnership on Rice Research (CORRA), where intellectual property and possible new laws in Asia for rice patenting were hot topics.

"These [intellectual property] legal requirements are new for many of us, but we have to study and introduce the necessary legislation as quickly as possible," says Joko Budianto, CORRA's chairman.

Research organizations that seek to further their intellectual-property rights were sharply criticized by groups such as the Spain-based Genetic Resources Action International (GRAIN).

"Pulling the rice cultures of Asia into the grips of corporate intellectual-property systems is not a win-win scenario," says GRAIN's director Henk Hobbelen. "Farmers will lose control over their production systems and biodiversity will suffer."

But CIMMYT trustee Fowler notes that research organizations have little choice in the rapidly evolving world of plant genomics. A year ago, he said, it would have been highly unlikely that the CIMMYT board would have



Crop protection: reforms will guarantee developing countries continued access to CIMMYT's seeds.

agreed to seek patent protection for its resources. But the centre fears private companies will exploit its openness and either patent a plant discovery or secure blocking patents that would limit its ability to improve crops.

As an example, the organization cites the case of a Colorado company that won US patent rights to a type of bean grown for years in Mexico and sold in California. The firm is suing in the US District Court to enforce its patent. Now the Mexican growers are alleging the patent is invalid.

Hope Shand, research director for the Canada-based Rural Advancement Foundation International, says these patent fights are exactly what organizations such as CIMMYT should avoid. "Filing patents and defending them is very expensive, taking scarce resources from agriculture research," she says.

Under a 1994 agreement, CIMMYT maintains designated germplasm in trust for the United Nations Food and Agricultural Organization (FAO) in Rome. CIMMYT's intellectual-property policy is designed to adhere to that agreement to provide designated germplasm to the international community, officials said.

For discoveries associated with non-FAO designed germplasm, the CIMMYT policy calls for analysis on a case-by-case basis. The policy will cover patents, licensing agreements and/or partnerships.

Timothy Reeves, CIMMYT's director general, says: "The whole rationale behind the intellectual property policy is to keep discoveries available to CIMMYT" in order to continue its mission of disseminating science for sustainable agriculture. **Rex Dalton**

... as Monsanto makes rice genome public

Paris & Sydney

The multinational life science company Monsanto this week announced the first complete sequencing of a crop plant genome — all 12 chromosomes of rice (*Oryza sativa*).

The company also promised that it would make the entire sequence available, with no strings attached, in the public domain, by donating it to the International Rice Genome Sequence Project (IRGSP), a Japanese-led consortium of ten publicly funded genome centres.

Takuji Sasaki, director of Japan's Rice Genome Research Program, the lead member of the consortium, told *Nature* that he applauded the Monsanto decision. He argued that the incorporation of Monsanto's database is an exemplary public-private sector

collaboration, similar to the joint completion of the *Drosophila* genome by Celera Genomics and the publicly funded *Drosophila* Genome Project.

The rice genome, at 400 Mb, is some 37 times smaller than wheat, and six times smaller than maize. But identifying the position of genes in rice will help in finding similar genes in these crops. The public rice project, which was launched in 1998, is being led by Japan with \$100 million through its Rice Genome Research Program. Ten million dollars is also being contributed each by China, the United States and India.

The US effort is publicly funded through the Institute for Genomic Research. Monsanto says that its data will be sent to IRGSP, and "as each segment of the sequence is completed, it will

be placed in the public domain in accordance with existing IRGSP policy."

The rice sequencing was carried out mainly by Leroy Hood at the University of Washington in Seattle, under contract for Monsanto. Sasaki says that the Monsanto sequence is of around 5X depth, and that many gaps remain. But merging the data with the 7X data from the public project should accelerate the completion of a 'finished' genome.

The IRGSP's initial target for completion was 2008 at a cost of \$200 million, but in a bid to shorten this Japan has increased its support for rice genomics three-fold, ironically in response to the threat that Celera and other companies might get their first (see *Nature* 401, 102; 1999).

Declan Butler & Peter Pockley

NASA pays the price of its dash for Mars

Washington

The coming reorganization of NASA's Mars programme in the wake of last year's double spacecraft disaster effectively ends a short but intense chapter in the saga of Red Planet exploration. The story began on 7 August 1996, when a team of researchers told a packed audience of news reporters that they had found evidence of fossil life in a Martian meteorite known as ALH84001 (see *Nature* **382**, 565; 1999).

Almost four years later, few people believe that ALH84001 really does contain fossils, and there is little scientific interest in pursuing the matter any further. The ambitious Mars exploration programme inspired by that announcement, which was to have culminated in bringing samples back to Earth in 2008, has followed roughly the same arc — initial excitement, reality check and a pause to regroup.

That reassessment includes a fundamental question — should sample return still be the main focus of the project? Before the 'Mars rock' hit, NASA had a plan for Martian exploration — a patient, methodical approach blessed by the National Academy of Sciences and other advisory bodies. Returning samples from Mars was a goal, but not the priority.

But publicity from the Martian meteorite kicked the programme into a higher gear, and the space agency soon committed itself to launching a Mars lander and orbiter at every 26-month opportunity and starting sample collection in 2003. NASA asked for, and Congress approved, a little more money for this accelerated programme, but not enough.

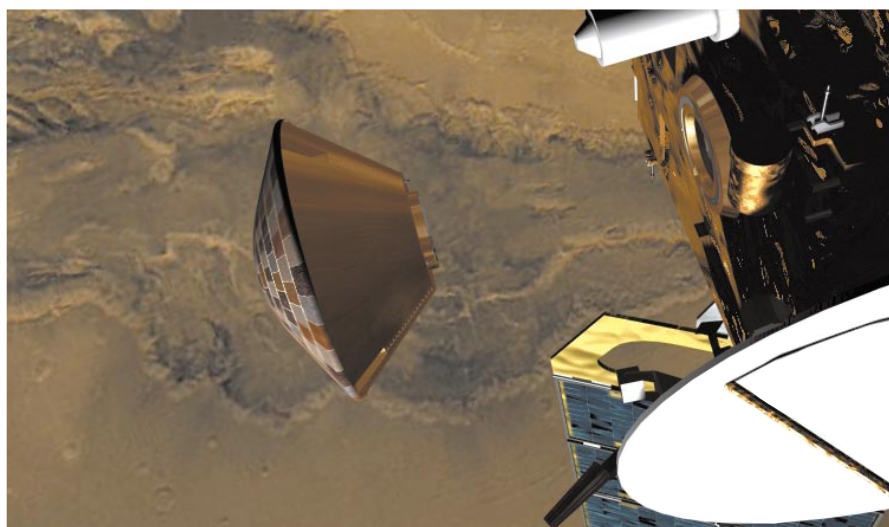
Two reports issued last week on the loss of the Mars Polar Lander and Mars Climate Orbiter (see www.jpl.nasa.gov/marsreports) make a convincing case that the projects were doomed from the outset by a lack of money, people and time.

Leisure programme

Now, as Cornell planetary scientist Steven Squyres, who is building the scientific payload for the next US Mars lander, says, "we have pushed the reset button". The 2001 lander is cancelled, and the 2003 sample return will almost certainly be delayed two years or more.

NASA science chief Ed Weiler now talks about a more leisurely ten-year programme instead of the enforced every-two-years pace. "Everything is on the table," says Squyres, as NASA sits down once again to reshape its Mars exploration "architecture", the next version of which is due this summer.

It is not even certain that the existing, worryingly complex plan for sample return — involving a US lander and rover, a small US rocket to lift samples off the surface, and a French orbiter to collect the samples and bring them back to Earth — will survive in its



Better luck? Britain's Beagle 2 is set to be the next craft to land on the surface of Mars.

current form. That undoubtedly is causing some anxiety at the French space agency CNES, which has signed up for a \$400 million role in the sample return mission.

But scientists may be relieved. Arizona State University planetary scientist Ronald Greeley, who chairs a NASA advisory committee on Mars exploration, says this hiatus gives us an opportunity to look at science goals from grass roots with all "artificial constraints" removed.

Landing problems

Most planetary scientists believe a Mars sample return is still called for eventually, but it will have to overcome several hurdles, starting with the landing on Mars. The last — and so far only — precise touchdown on the planet's surface were by the Viking spacecraft in 1976. Following the Mars Polar Lander's crash, says Squyres, "we don't have a validated way to land, other than the Mars Pathfinder airbags". Engineers considering the next generation of lander will have to choose between airbags, legs or some other mechanism such as crushable skirts.

NASA will also need to develop a means of avoiding very rough landing sites, which may require that a small, cheap lander, airplane or other scout be sent to investigate the site beforehand — a conclusion reached by a NASA-sponsored sample return workshop held several months before the Mars rock accelerated the programme.

If the driving force behind a sample return mission is the search for evidence of past or present life, scientists will have to develop clear "biomarkers" that can unambiguously settle the question of whether a returned sample contains evidence of biology. Those biomarkers do not yet exist, which partly accounts for some scientists' reluctance to rush headlong into sample return. According

to Greeley, the research community is divided between those who believe any returned rock would teach us much about Mars, and those who think we should wait until we know what we are searching for.

"Sample return has to take place in the context of a broader understanding of the planet," says Squyres. "There's clearly a place for both sample return and *in-situ* science. The real questions is, in which order does it make the most sense to do things."

That matter has already been settled by NASA's abdication of the 2001 landing opportunity. The next spacecraft firmly scheduled to touch down on Mars — if private funding comes through to supplement the British government's \$13 million investment — is the tiny Beagle 2 lander that will ride along with the European Space Agency's Mars Express orbiter to be launched in 2003.

The \$40 million lander, the brain child of Colin Pillinger of the Open University, has an exobiological focus but modest scientific goals, according to Andre Brack of the Centre de Biophysique Moléculaire in Orleans, France, who chairs the science team for Beagle 2.

The lander, which will use Pathfinder-style airbags to cushion its fall to the surface, will be equipped with a robot arm and a 'mole' for digging a few tens of centimetres below the surface. The main objective, says Brack, is simply to find organic material underneath the topmost layer of soil.

Viking failed to find organic material on the surface, and effectively killed interest in returning to Mars for two decades, says Brack. If Beagle 2 could pick up where Viking left off, it would be a great contribution. Not fossils perhaps, but the beginning of a new and more patient effort to understand Mars.

Tony Reichhardt

Royal Society delays conference on origins of HIV

London Britain's Royal Society has delayed a controversial meeting to discuss the origins of HIV that was to have taken place in London next month (see *Nature* 404, 9; 2000). The meeting's organizers say that this is partly to allow the publication of pertinent scientific evidence.

The meeting will now take place in September. Ed Hooper, author of the book *The River* (Allen Lane), in which the idea that HIV could have originated in a contaminated batch of polio vaccine is put forward, claimed that the delay is "an attempt to load the dice".

But in a letter to *The Guardian* replying to a report on the shift of dates, Sir Aaron Klug, the president of the Royal Society, wrote that "because of the dissension aroused by [this] hypothesis, we found it impossible to arrive at a well-balanced set of speakers in time for the meeting in May".

US AIDS chief decides to spend more time with family

Washington The director of the US Office of AIDS Research (OAR) at the National Institutes of Health (NIH) is stepping down after

two years in the post. Neal Nathanson announced last week that he will return on 1 September to the University of Pennsylvania Medical Center, where he was formerly the vice-dean for research and research training.

Nathanson, 72, a virologist, said he was leaving the post to spend more time with his family. He has split his time between Philadelphia and the NIH while at the OAR. Gregg Gonsalves, policy director at the Treatment Action Group, a national AIDS research advocacy organization, called Nathanson's departure "a big surprise" and a considerable loss.

Stanford woman wins discrimination case

San Diego A jury in San Francisco last week ruled that Stanford University retaliated "maliciously" against a former computer scientist who complained about alleged sex discrimination, awarding the woman \$541,000 in damages. Colleen Crangle claimed she was forced out of the university by male superiors after she objected to being subordinated to a male researcher.

Stanford says it plans to appeal against the verdict, which its attorneys called disappointing. Crangle is one of 32 former and current Stanford employees whose complaints have prompted an investigation

by the US Department of Labor into allegations of job discrimination. Stanford officials deny the allegations.

South African government tight-lipped over AIDS

Cape Town South Africa's health minister, Manto Tshabalala-Msimang, last week denied reports that the country's president, Thabo Mbeki, believed that HIV does not cause AIDS. But she refused to say whether she personally held this belief.

Tshabalala-Msimang's statement followed Mbeki's response to calls for him to reconsider the government's policy of not providing antiretroviral drugs to HIV-positive pregnant women. Mbeki replied that he would continue to question conventional wisdom on AIDS (see *Nature* 402, 3; 1999).

Germany and Israel in 'science for peace' bid

Munich The research ministers of Germany and Israel have invited scientists in Palestine, Egypt and Jordan to increase their participation in publicly funded German-Israeli research projects as a way of supporting the peace process in the Middle East.

Areas such as water management, marine research and environmental sciences would

be the first choice for extended scientific cooperation, said the ministers, Edelgard Bulmahn and Matan Vilnai, at a celebration last month in Berlin of the twenty-fifth anniversary of the start of scientific collaboration between their countries. The federal German science ministry has funded more than 500 joint German–Israeli research projects since 1975.

European academies join calls for genome access

Prague Thirty-five European academies of science have endorsed a plea by the presidents of the US National Academy of Sciences and the Royal Society of London to respect the need for unimpeded access to raw human genome sequence data (see *Nature* 404, 325; 2000). The endorsement took the form of a statement agreed at the general assembly of the All European Academies in Prague last week.

The statement says it is “critical” that the development of medical applications of the information in the genome sequence is “not impeded by any restrictions on access to the raw material of the genome sequence”. Attempts to claim patents on raw segments of the sequence, it adds, “disregard the extent to which publicly funded research has contributed to the progress that has been achieved”.

Chemist declines top post in Taiwan

Tokyo To the disappointment of most Taiwanese voters — and newly elected president, Chen Shui-bian — Lee Yuan-tseh, the 1986 Nobel prizewinner in chemistry, last week declined Chen’s offer to become prime minister. Lee will instead return as president of the Academia Sinica, a post he vacated a week before the presidential election to support Chen (see *Nature* 404, 430; 2000). He will also act as a government adviser on education and relations with China.

Lee’s decision came as a relief to members of the academy, who declined officially to accept Lee’s resignation and launched a campaign to dissuade him from leaving the institution. Yang Kuo-su, vice-president of the Academia Sinica, said recently that the academy needed “proper leadership” to prevent Taiwan’s research system from suffering “degradation”.

Correction

In the news item “Global warming sceptics left out in the cold” (see *Nature* 403, 233; 2000), it was stated erroneously in the first paragraph that “the Earth’s atmosphere is warming up”. The article should have stated “surface” and not “atmosphere”. This error did not affect the rest of the news report.



Biotechnology meeting attracts protests in Boston

Boston More than 2,500 protesters, some dressed as mutant plants and animals, expressed their concerns over the spread of biotechnology at the BIO 2000 conference in Boston last week (above), attended by 8,000 biotech researchers and business leaders.

STEPHAN SAVOIA/AP

Whatever happened to leptin?

Just five years ago, it seemed that a single protein might reverse the rising tide of obesity. What worked for mice has not yet translated to people. But watch this space, says Marina Chicurel.

By any reckoning, US\$20 million is a lot to spend on a protein. But in May 1995, Amgen of Thousand Oaks, California, paid just that for the commercial rights to leptin, a hormone that made fat mice slim. Identified only six months earlier, leptin could be injected into grotesquely obese, leptin-deficient mice where it curbed their voracious appetites and boosted their metabolic rates. Within a month, and with no apparent side-effects, these mice lost almost half of their excess weight.

Leptin was hailed as the wonder drug to defeat a burgeoning epidemic of obesity. But five years down the line, this simple script has had to be rewritten. First, data from studies on rodents and humans hinted at leptin's therapeutic shortcomings. Then, last October, came the news that the protein had performed poorly in its first clinical trial¹.

Although Amgen may not have its hands on a pharmaceutical blockbuster, obesity researchers are enthused by the wealth of advances that have followed leptin's discovery. And some of these advances may, in the long term, fulfil the therapeutic hopes once pinned on leptin. "It transformed the world of obesity from a backwater of totally ignored, rather dodgy research into one of the primary areas of biomedical research," says Stephen O'Rahilly of the University of Cambridge.



Weight watcher: Friedman's discovery kick-started a vibrant field of research.

Leptin is significant because it plays an important role in energy homeostasis — the physiological balancing act that regulates our food intake and our storage and expenditure of energy. It also seems to affect other physiological systems, including carbohydrate metabolism, reproduction and bone formation. "As a hormone, leptin is probably as important as insulin," says Joel Elmquist of the Beth Israel Deaconess Medical Center and Harvard Medical School in Boston.

Exercising control

Leptin's fundamental power stems from its ability to keep the brain informed about the status of the body's energy stores. The amount of leptin secreted by fat cells depends on their size. When food is scarce, the cells shrink and secrete less leptin into the bloodstream. This reduces the activity of nerve cells that have leptin receptors and are found in a region of the brain called the hypothalamus. In turn, this decrease in neural activity has two effects on other neurons in the hypothalamus: it boosts the release of neuropeptide Y, a neurotransmitter that increases appetite, and it suppresses α -melanocyte-stimulating hormone (α -MSH), which blocks feelings of hunger. Through these pathways, and perhaps through direct effects on organs such as the pancreas and liver, leptin ensures energy homeostasis.

But the balance sometimes breaks down, probably because leptin and its associated neural circuits evolved to maintain adequate fat reserves, rather than to avoid excess fat storage — our distant ancestors routinely had to cope with food scarcity.

This view of leptin is different from the protein's initial characterization as an 'anti-obesity' hormone. But given the circumstances of leptin's discovery, it is easy to see why the early view prevailed. Since 1950, researchers had realized that the bizarre plight of the *ob/ob* mouse was the result of a genetic abnormality. This inbred strain eats almost continuously, weighs three-times as much as normal mice, and suffers from diabetes. But it was not until 1994 that Jeffrey Friedman and his colleagues at the Howard Hughes Medical Institute at Rockefeller Uni-



Despite its disappointing performance so far, leptin is still the focus of interest from drug developers.

versity in New York sequenced the *ob* gene, identified its human counterpart, *Ob*, and characterized the protein they produce^{2,3}. The team called the protein leptin, from the Greek word *leptos*, or thin.

Within months, several teams, including Friedman's, had injected leptin into *ob/ob* mice and transformed them into healthy animals with normal body weights³⁻⁵. Leptin specifically reduced fat mass, sparing other tissues. "Never had there been any clear circulating factor that was involved in body weight regulation," says Steven Heymsfield

of St Luke's-Roosevelt Hospital in New York, who has collaborated with Amgen on human trials of leptin. "So that was a major, major breakthrough."

Predictions of potency

Friedman believes it was the powerful images of the transformed obese mice that ignited enthusiasm for leptin. Given the extreme competitiveness of the pharmaceutical industry, Amgen was quick to buy the rights from Rockefeller University.

Almost immediately, however, studies on rodents and people began to reveal that there was much more to leptin than first met the eye. The hormone's precise physiological role seems to vary from species to species. O'Rahilly's group in Cambridge, for instance, identified two obese, leptin-deficient children — the human equivalent of *ob/ob* mice⁶. But, unlike the mice, the twins did not suffer from other conditions such as hypothermia and severe diabetes. Nevertheless, when O'Rahilly's group injected one child with leptin, she lost weight rapidly⁷.

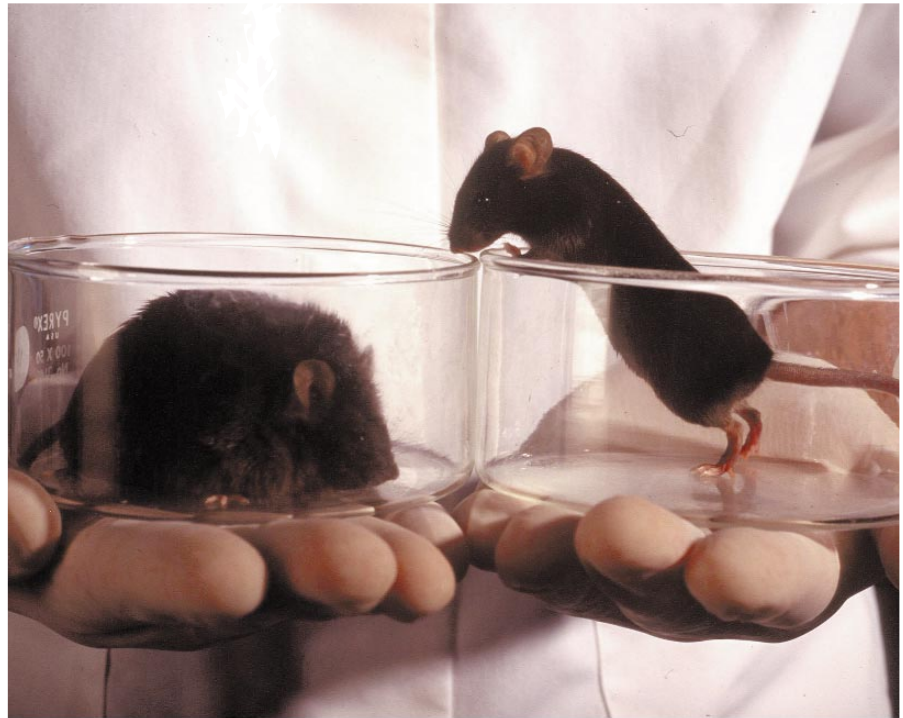
Researchers also soon realized that most obese people are not deficient in leptin. In fact, the majority have higher than normal levels of the hormone^{8,9}. Many obese people seem to be, at least to some extent, resistant to leptin. In addition, leptin levels vary greatly between people with identical percentages of body fat, indicating that a variety of factors, genetic and environmental, might contribute to people's sensitivity to the hormone^{8,10}.

Despite this evidence, Amgen remained hopeful. After all, diabetics with high circulating levels of insulin can sometimes benefit from receiving even more of the hormone. But the limits to leptin's therapeutic potential became clear last year with the publication of results from an Amgen trial in which 73 obese volunteers injected themselves with leptin or a placebo¹.

Of the 47 patients who completed the study — many withdrew because of problems with the injections — the eight receiving the highest dose lost an average of 7.1 kg, while the twelve taking a placebo lost 1.3 kg. The effects varied widely among patients — from a loss of about 15 kg to a gain of 5 kg within the group treated with the highest dose. These doses also induced skin irritation and swelling at the injection site. "The data reported were fairly moderate," concedes Carl LeBel, associate director of product development at Amgen.

But these setbacks are not the end of the road for leptin. The protein's discovery has provided countless new leads for drug development. Within the past five years, researchers have identified more than a dozen key components in the neurocircuitry underlying appetite control. They have also begun to tease out the intracellular signalling pathways activated by leptin.

Several companies are exploring the therapeutic possibilities of molecules that act



Before and after: dramatic images such as this one fired pharmaceutical enthusiasm for leptin.

at later stages in the process, after leptin has done its work. Millennium Pharmaceuticals in Cambridge, Massachusetts, for example, is working with drugs giant Hoffmann-La Roche to seek ways of heightening the effects of the appetite-suppressing α -MSH. This neurotransmitter acts on melanocortin-4 (MC-4) receptors, found on cells in the hypothalamus that are connected to leptin-responsive neurons. Mice lacking MC-4 receptors or pro-opiomelanocortin, the precursor of α -MSH, are obese¹¹. Data from two unpublished studies, including one from O'Rahilly's group, suggest that up to five per cent of obese children carry mutations in the melanocortin system, making these mutations the most common genetic defects to be associated with human obesity to date.

But boosting a protein's function is difficult, says Eleftheria Maratos-Flier of the Joslin Diabetes Center and Harvard Medical School. So other companies are focusing on targets that have weight-reducing effects when their function is blocked. Some researchers are trying to block melanin-concentrating hormone (MCH). Although there is disagreement about MCH's exact position in the circuitry controlled by leptin, mice that lack the hormone are much leaner than normal¹², and if the hormone is injected

into rats' brains, the animals immediately start eating excessively¹³.

Peering into the neurons that sport leptin receptors, other groups are looking for intracellular drug targets. Louis Tartaglia at Millennium believes the most promising approach is boosting the responsiveness of the signalling pathway activated by leptin. One component of this pathway, called suppressor of cytokine signalling-3 (SOCS-3), was recently identified in Jeffrey Flier's lab at Beth Israel Deaconess and Harvard¹⁴. Leptin seems to boost production of SOCS-3, which in turn seems to stifle further leptin signalling. It is possible, therefore, that the resistance of obese people to leptin is the result of SOCS-3 overactivity.

Copycat molecules

Dissection of leptin's signalling path has also helped reveal the weight-regulating properties of another molecule — ciliary neurotrophic factor (CNTF). This is one of a range of nerve growth factors being explored as treatments for neurodegenerative diseases. One trial using CNTF to treat amyotrophic lateral sclerosis revealed its leptin-like behaviour in humans. The patients did not show much neurological improvement, but they lost lots of weight¹⁵.

The cell-surface receptor for CNTF closely resembles the leptin receptor, it activates similar intracellular signalling pathways, and it is found in related areas of the hypothalamus. Extending work on mice, in which Ralph Lauffer's team at IRBM, the Institute for Research in Molecular Biology in Rome, showed that CNTF made obese mice lose fat¹⁶, George Yancopoulos and colleagues at Regen-

**nature
insight**

Turn to page 631
for the first in a
regular series of
topical reviews.

Obesity

eron Pharmaceuticals in Tarrytown, New York, started experimenting with axokine, a patented derivative of CNTF. When they made normal rodents obese by putting them on a high-fat diet, the animals became resistant to leptin. But when the researchers injected these same animals with axokine, the rodents shed 30 per cent of their body weight and maintained their moderate eating habits, even several days after their last injection.

Sights set on success

Last September, Regeneron completed a phase I clinical trial of axokine. Although at high doses the drug seemed to cause cold sores and nausea, patients tolerated low doses well. The results have yet to be published, but most patients lost significant amounts of weight, reducing their energy intake by about 500 calories a day. Yancopoulos calls axokine "the most promising weight loss agent ever described", but others are more cautious. Tartaglia at Millennium, for example, worries that tinkering with a neuronal growth factor could have delayed effects that are difficult to predict.

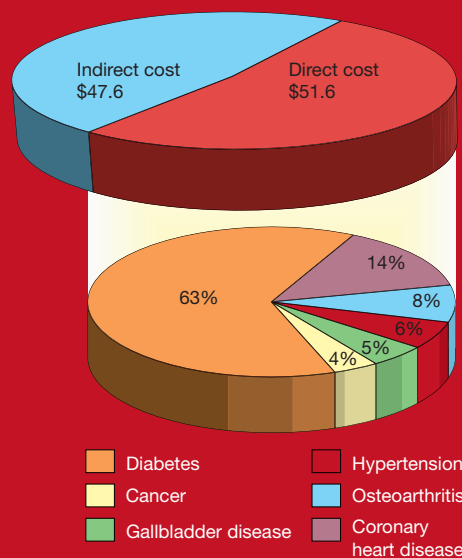
Despite its disappointing performance so far, leptin itself is still the focus of interest from drug developers. Many scientists, including those from Amgen, think leptin might help people maintain weight loss. For most people, the hardest part of dieting is not shedding weight, but keeping it off — they tend to return to their previous weight. The drop in the level of circulating leptin is probably partly to blame for this because the ensuing shrinkage of fat cells sounds a 'starvation alarm' that boosts the appetite. Leptin injections could help overcome this by fooling the brain into thinking that its energy stores are not at risk.

Amgen is also designing second-generation versions of leptin. The goal is to make leptin mimics that are more soluble — to increase the dose that can be injected into patients — and more stable, to last longer in the bloodstream. Other researchers are designing small molecules with leptin-like activity that can easily cross the blood-brain



Counting the cost of obesity

For the first company to bring a potent weight-loss drug to the market, the rewards could be huge. Across the industrialized world, the prevalence of obesity is rising. Last month, the Worldwatch Institute in Washington DC reported that 1.1 billion people worldwide are now suffering from the health consequences of overconsumption — equal to the number plagued by hunger. In the United States alone, where more than one in three adults is classified as obese, the condition is estimated to cause 300,000 premature deaths each year. And, as these figures reveal, the associated burden of disease causes a significant annual dent in the US economy — both from the direct cost of treating conditions triggered by obesity, and from lost economic productivity.



Source: Wolf, A. M. & Colditz, G. A., *Obes. Res.* 6, 97-106 (1998)

barrier. This could be important, because the ratio of leptin in cerebrospinal fluid to that in blood is four times lower in obese people than in those of normal weight^{17,18}.

The success or failure of these strategies, however, depends on how leptin sensitivity varies from person to person. Because of their resistance to leptin, flushing the cerebrospinal fluid of obese people with leptin-like molecules might make very little difference. So Amgen is searching for biomarkers that correlate with leptin sensitivity. This may allow leptin to be prescribed selectively to those who are likely to respond more strongly to the hormone.

Leptin's reach may also extend to the treatment of other, related diseases. Roger Unger at the University of Texas Southwestern Medical Center in Dallas believes that obesity-induced

diabetes develops when pancreatic β -cells die after accumulating excessive fat in their cytoplasm. By stimulating receptors on the surface of β -cells, Unger thinks leptin can stop this from happening. But other researchers caution that leptin's interaction with insulin remains mysterious and controversial.

While Amgen's multimillion-dollar bet on leptin doesn't look like yielding an immediate bonanza for the company, obesity researchers agree that the protein has kick-started an exciting field — both for basic research and for drug development. "Whatever has happened to leptin as an immediate therapeutic tool, as a discovery that sparked an area of biomedical research, it's phenomenal," says O'Rahilly.

Marina Chicurel is a freelance writer in Santa Cruz.

- Heymsfield, S. B. *et al.* *J. Am. Med. Assoc.* **282**, 1568-1575 (1999).
- Zhang, Y. *et al.* *Nature* **372**, 425-432 (1994).
- Halaas, J. L. *et al.* *Science* **269**, 543-546 (1995).
- Pelleymounter, M. A. *et al.* *Science* **269**, 540-543 (1995).
- Weigle, D. S. *et al.* *J. Clin. Invest.* **96**, 2065-2070 (1995).
- Montague, C. T. *et al.* *Nature* **387**, 903-908 (1997).
- Farooqi, I. S. *et al.* *N. Engl. J. Med.* **341**, 879-884 (1999).
- Maffei, M. *et al.* *Nature Med.* **1**, 1155-1161 (1995).
- Considine, R. V. *et al.* *N. Engl. J. Med.* **334**, 292-295 (1996).
- Lonnqvist, F., Arner, P., Nordfors, L. & Schalling, M. *Nature Med.* **1**, 950-953 (1995).
- Huszar, D., *et al.* *Cell* **88**, 131-141 (1997).
- Shimada, M., Tritos, N. A., Lowell, B. B., Flier, J. S. & Maratos-Flier, E. *Nature* **396**, 670-674 (1998).
- Qu, D. *et al.* *Nature* **380**, 243-247 (1996).
- Bjorbaek, C., Elmquist, J. K., Frantz, J. D., Shoelson, S. E. & Flier, J. S. *Mol. Cell.* **1**, 619-625 (1998).
- ALS CNTF Treatment Study Group *Neurology* **46**, 1244-1249 (1996).
- Gloaguen, I. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 6456-6461 (1997).
- Caro, J. *et al.* *Lancet* **348**, 159-161 (1996).
- Schwartz, M. W., Peskind, E., Raskind, M., Boyko, E. J. & Porte, D. Jr *Nature Med.* **2**, 589-591 (1996).

Diet plan: Amgen still believes leptin is the key to the lucrative obesity market.



ERIC SANDER/LIAISON AGENCY

Assessment mismatches must be sorted out: they leave species at risk

Sir—Although Rodríguez *et al.* are correct that nationally endemic taxa should be classified identically on national and global Red Lists¹, they provide no evidence to support their inferences that national Red Lists “are more accurate” than global Red Lists, or that global lists ignore a “wealth of local data”. Moreover, we see certain problems with the national Red Lists they analyse which suggest these inferences may be mistaken.

Only nine of the 70 different assessments tabulated by Rodríguez *et al.* are due to different information about species. Thirty-two result from differences between IUCN assessors and the national assessors over the choices of taxa for consideration—especially taxa from groups such as reptiles, amphibians, bony fishes and invertebrates, which IUCN has only partially considered²—and the inclusion of sub-species, which have not been a major focus of the IUCN Red List. Seven species were listed as Data Deficient on the national lists, whereas for the global list the same information was considered adequate to place them in a threatened category. Of particular concern are the cases resulting from inconsistent use in the national lists of the IUCN Red List Criteria³ (see Table 1).

Many of these inconsistencies are compounded by national Red List assessments not making their use of the Red List Criteria explicit for each taxon, as required³, and/or failing to provide supporting information (despite the claim that national Red Lists utilize a “wealth of local data”, the Argentine⁴ and Ecuador⁵ lists contain no supporting data whatsoever).

At present IUCN recommends that experts making assessments for the global Red List should consult national authorities in reaching a decision. IUCN would like to move rapidly towards a situation where the national assessments could be used without any intermediate steps to ensure

standardization of approaches. To this end IUCN has established two key initiatives.

First, to improve information flow and quality, IUCN is appointing Red List Authorities (RLAs) who will ensure that all assessments are done in a fully consultative manner and are well documented and peer reviewed (a model established by BirdLife International^{6,7}). The RLAs will have access to shared databases of species information that will be based in part on interactive web technologies (see <http://www.iucn.org/themes/ssc/programs/sis.htm>).

Second, IUCN aims to improve consistency in the use of the Red List Criteria. It is running a series of regional Red List workshops around the world: the first, held in Sri Lanka in September 1999, involved participants from 12 Asian countries. Other workshops are planned for Mesoamerica, South America, southern Africa and east Africa. IUCN is also recommending and distributing standard computer software⁸ that takes assessors systematically through the categorization process. Progress in all this is constrained only by resources.

Craig Hilton-Taylor*, **Georgina M. Mace†**, **David R. Capper‡**, **Nigel J. Collar‡**, **Simon N. Stuart§**, **Colin J. Bibby‡**, **Caroline Pollock***, **Jørgen B. Thomsen||**

*IUCN/SSC UK Office, 219c Huntingdon Road, Cambridge CB3 0DL, UK

†Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

‡BirdLife International, Wellbrook Court, Girton Road, Cambridge CB3 0NA, UK

§IUCN/SSC, Rue Mauverney 28, Gland CH-1196, Switzerland

||Conservation International, 2501 M Street, NW, Suite 200, Washington, DC 20037, USA

1. *Nature* **403**, 241 (2000).
2. Baillie, J. & Groombridge, B. (eds) 1996 *IUCN Red List of Threatened Animals* (IUCN, Gland, 1996).
3. IUCN Species Survival Commission IUCN Red List Categories (IUCN, Gland, 1994).
4. García Fernández, J. J., et al. *Libro Rojo de Mamíferos y Aves Amenazados de la Argentina* (Fundación para la Conservación de las Especies y el Medio Ambiente, Buenos Aires, 1997).
5. Granizo, T. M., et al. *Lista de Aves Amenazadas de Extinción en el Ecuador* 1–31 (IUCN-Sur, CECIA, INEFAN, EcoCiencia y BirdLife International, Quito, 1997).
6. Collar, N. J. et al. *Threatened Birds of the Americas: the ICBP/IUCN Red Data Book* (International Council for Bird Preservation, Cambridge, 1992).
7. Collar, N. J., Crosby, M. J. & Stattersfield, A. J. *Birds to Watch* 2:

Table 1 Reasons for disparities between threatened endemic animal taxa (Critically Endangered, Endangered and Vulnerable) on the IUCN global Red List and on national Red Lists

	Taxa					
	Mam	Birds	Rep	Amph	BF	All
Number of threatened taxa on both lists	8	30	1		1	40
Species not assessed by IUCN	1		1	7	7	16
Species on IUCN but not national lists	4				3	7
Sub-specific taxa not assessed by IUCN		4	1			7
Sub-specific taxa on IUCN but not national lists	2					2
Different use of categories on national lists	2	4	1			7
Different use of criteria on national lists	7	15				22
Different information used for listings	3	6				9
Total number of threatened taxa	29	59	4	7	11	110

Disparities are those between the IUCN global Red List² and national Red Lists of Argentina⁴, Bolivia⁵, Ecuador⁵ and Venezuela¹⁰.

Key: Mam=Mammals, Rep=Reptiles, Amph=Amphibians, BF=Bony Fish, All=All taxa combined. No disparities were found for Crustaceans or Insects.

The World List of Threatened Birds (BirdLife International, Cambridge, 1994).

8. Akçakaya, H.R. & Ferson, S. *RAMAS Red List: Threatened Species Classifications Under Uncertainty*, Version 1.0 (Applied Biomathematics, New York, 1999).
9. Ergueta, P. & de Morales, C. (eds) *Libro Rojo de los Vertebrados de Bolivia* (Centro de Datos para la Conservación, La Paz, 1996).
10. Rodríguez, J. P. & Rojas-Suárez, F. *Libro Rojo de la Fauna Venezolana* 2nd edn (PROVITA, Fundación Polar, Wildlife Conservation Society, PROFAUNA-MARNR, UICN, Caracas, 1995).

Fooled, but not foolish

Sir—References to *National Geographic* in your Opinion article “Fossil Smuggling Unopposed”¹ are false. The article on new dinosaurs from China was not “naively and hastily” published. It had been in preparation for a full year before publication and during this time was reviewed by six leading paleontologists and ornithologists.

Pertinent information concerning the integrity of the specimen known as *Archaeoraptor* was not made available to us or to them before publication. *National Geographic* is sparing no effort to investigate all issues raised concerning this specimen.

I also take strong exception to the anonymous quote in your editorial. That statement, which implied that our reporting was biased, was made by a person who is known to be a vocal opponent of the dinosaur–bird theory.

There is no question that a major problem in the field of palaeontology is the commercial trade in fossils. *National Geographic*, for its part, has a policy of not photographing or writing about fossils that are of illegal status, and we did not consider including *Archaeoraptor* in the article until assured it would be returned to China. We are proud of *National Geographic*’s role in that process.

William L. Allen

National Geographic Magazine, 1145 17th Street, Washington, DC 20036

1. *Nature* **403**, 687 (2000).

How much use is the Human Genome Project?

Sir—Current hype about the expected completion of the Human Genome Project demands some clarification of the details as well as of the wider scientific implications. Although initially conceptualized more broadly, the project is effectively about determining the sequence of bases in the human genome. This is not the same as trying to understand the program that is encoded in human DNA, an obvious requirement for any pharmacological application. Consequently, the results will be in the merely descriptive naturalistic

▶ tradition (along the lines of 'the mean content of bases G and C of sequence so-and-so is 47.8 per cent'). Technical development has always had that effect on scientific disciplines, for example the electron microscope, the radio telescope or the automated DNA sequencer.

Of course, researchers are always quick to emphasize the importance of their work to whatever application is in vogue, and curing diseases is a worthy goal. But how specifically will the Human Genome Project help to achieve this end? A look at any gene (as opposed to a sequence) map from any species reveals what looks like an explosion in a slaughterhouse. Where is the order we need, to make sensible rather than trial-and-error genetic manipulations?

Should scientists' claims of applicability for their results be acknowledged as a mechanism to secure funding rather than having any realistic basis? 'Science is a process and not a series of final states' is the somewhat trite argument to justify goals not achieved. A series of simple descriptive, but highly technical, publications ensures that research money will be channelled into well-trodden paths in the future.

In any case, pharmacogenomics requires an understanding of the apparent genetic 'disorder' in any organism's genome, of genotype-phenotype mapping, of gene-gene interactions (epistasis), of intraspecific genetic variability, and of self-organizational processes, rather than endless lists of DNA bases.

Sol Hadden

*Advisory Council for Advanced Concepts,
206 Rhodes Hall, Cornell University,
Ithaca, NY 14853, USA*

Genes: we can't expect full understanding yet

Sir—The editing of one phrase in our statement about ownership of the human genome¹ might lead readers to misconstrue our intention. In the sentence printed as "The intention of some university and commercial interests to patent the DNA sequences themselves, thereby staking claim to large numbers of human genes without necessarily having a full understanding of their functioning, strikes us as contrary to the essence of patent law", the word "full" should have been deleted.

In fact, the full understanding of a gene is likely to take many decades to accomplish, and such a criterion would clearly be unreasonable with respect to what is patentable. Our point is that some level of understanding of specific function is important before a patent is awarded, as this is a necessary precursor to the claim of a substantial utility.

We also wish to comment that it is not the case, as implied in the Opinion article in the same issue² that the main target of our statement was Celera Genomics. We were addressing important issues of broad public policy, and our focus was primarily the patent offices and the law courts, in which the validity of claims for gene-sequence patents will be decided.

Bruce Alberts*, Aaron Klug†

**National Academy of Sciences, 2101 Constitution Avenue NW, Washington DC 20418, USA*

†Royal Society, Carlton House Terrace, London SW1Y 5AG, UK

1. *Nature* **404**, 305 (2000).

2. *Nature* **404**, 317 (2000).

Garlic study vindicated by official investigation

Sir—Your News story "German garlic study under scrutiny"¹ reports allegations of data manipulation and incorrect data analysis raised in a German newspaper about a study we carried out using a garlic preparation called Kwai. They were made after the results of our clinical study "Randomized placebo-controlled double-blind study on antiatherosclerotic effect of Kwai in common carotid arteries and femoral arteries" were published².

As mentioned in your story, an official committee at Humboldt University was set up to investigate the claims of falsification. It has now announced that the accusation of data manipulation is unfounded.

The investigating committee has found that the clinical trial had been sanctioned in advance by the relevant ethics committee and that the patients had agreed individually, in writing, to participate. The use of an alternative statistical evaluation was explained in detail in our original article², but we have clarified this matter in a letter to the same journal³.

We have also confirmed the plaque reduction reported in our published article² by two more ultrasound photos of the same patient from the verum group (with initials and date at the same examination time) using the sector scanner, a different ultrasound system (data not shown). We have examined and confirmed the reproducibility of our earlier ultrasound photos, and we will present our confirmatory findings in a future article.

J. Koscielny, R. Schmitt, H. Radtke, R. Latza, H. Kiesewetter

Institute for Transfusion Medicine, Medical Faculty Charité, Humboldt University, Campus Charité Mitte, Schumannstr. 20/21, D-10117 Berlin, Germany

1. *Nature* **401**, 629 (1999).

2. Koscielny, J. et al. *Atherosclerosis* **144**, 237–249 (1999).

3. Siegel, G. *Atherosclerosis* **148** (in the press).

Learn lateral thinking first and specialize later

Sir—Frederick Seitz's Millennium Essay, "Decline of the generalist"¹, voices an important truth about narrow specialization in science. The twentieth century was the age of scientific specialization, and in many areas this will continue.

However, the pendulum begins to swing back. *Nature* readers can hardly have overlooked those systems that today challenge science and society: geophysiological systems of ocean, weather and global warming; biophysical-mathematical systems such as brain function and animal behaviour; and other natural systems, large and small. All of these trample briskly across traditional scientific boundaries. So is there not a new urgency about the inter- and multidisciplinary teaching of science? Should we not prepare science graduates whose careers will take other directions to understand something of the science and technology that will dominate their world?

General and specialist aims in scientific education are not incompatible. The answer is to offer university science courses broad enough to encourage lateral thinking across two or more disciplines, while positioning the graduate to embark upon specialization in one area. Additional specialist training can readily be acquired later, but the habit of lateral thought cannot—it must come first. Only some students will wish to follow this track into science, which makes its own special intellectual demands, but they will be better for doing so.

Nicholas J. Kuhn

School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK

1. *Nature* **403**, 483 (2000).

Religion has its place but don't pretend it's science

Sir—I protest most strongly at Geoffrey Cantor's statement in his Millennium Essay¹ that I have made tirades against religion and that I regard it as the enemy of science. This is simply false; see what I have written on this topic, as in my *Unnatural Nature of Science* (Faber, London, 1993).

I do, however, follow David Hume, who made clear that religion is based on faith, science on reason. I only oppose religion in relation to science when people make scientific claims for it, for example in supporting creationism.

Lewis Wolpert

Department of Anatomy and Developmental Biology, University College London, Gower Street, London WC1E 6BT, UK

1. *Nature* **403**, 831 (2000).

Tracking down a lethal inhalation

How a group of investigators unearthed the truth behind a Russian epidemic.

Anthrax: The Investigation of a Deadly Outbreak

by Jeanne Guillemin

University of California Press: 1999. 321 pp.
\$27.50, £17.50

Ed Regis

In the first week of April 1979, an anthrax epidemic broke out in Sverdlovsk, a city of 1.2 million situated in the Ural Mountains, some 900 miles east of Moscow. For the next 15 years, those were practically the only facts about the case that everyone could agree upon — everything else was a mass of conflicting claims and uncertainty. Some news reports spoke of as many as 1,000 dead, whereas the Soviets themselves were admitting to only 64 fatalities. There were rumours that the outbreak was caused by an explosion in a biological weapons research facility, but the Russians denied them (the Soviet Union was, after all, a signatory to the 1972 Biological Weapons Convention prohibiting such research). Instead, they attributed the disease, which they called 'Siberian ulcer', to people having eaten anthrax-contaminated meat.

The mystery only deepened as additional facts emerged from the case: the developmental curve of the outbreak described what looked like a prolonged epidemic spread across six weeks, whereas an outbreak resulting from an explosive release of airborne anthrax spores would be expected to last only a short time, with virtually all of the cases clustered at the beginning. All the victims were adults; there were no children or infant deaths. This was hard to explain on either the contaminated-meat or sudden-explosion scenario. (A third scenario, which combined features of each of the others, was that an anthrax aerosol had been created by the burning of infected animal carcasses.)

In 1988, to quell the rumours, and to put to rest the lurking suspicions that the Soviet Union had violated the treaty they had signed, two Soviet physicians who had been involved in the outbreak came to the United States and made a case for the tainted-meat scenario before a number of scientific groups. Despite the absence of positive clinical and epidemiological evidence, many in the audience, including Jeanne Guillemin, were substantially convinced by their presentations.

In 1991, the *Wall Street Journal* ran a series of articles by Peter Gumbel, Moscow bureau chief, casting doubt on the official explanation. For one thing, the meat-processing plant alleged to have been the source of the anthrax did not exist. For another, many of the victims lived or worked near the secret military facility in Sverdlovsk.

Finally, in 1992, Harvard biologist Matthew Meselson and his wife Guillemin, a Boston College sociologist, with a team of collaborators including a veterinarian and a pathologist, made a private, fact-finding expedition to Sverdlovsk to learn the true cause of the outbreak. They examined the available medical evidence and unearthed vast amounts of new information. Some of the evidence was of dubious quality or ambiguous: the investigators were shown slides purporting to be of samples taken from the victims, but there was no way of knowing whether the slides had been faked, or had originated from some other epidemic. Moreover, since there had been so few cases of inhalational anthrax in history, nobody was sure exactly what pathological signs would differentiate gastrointestinal from pulmonary anthrax during autopsy procedures.

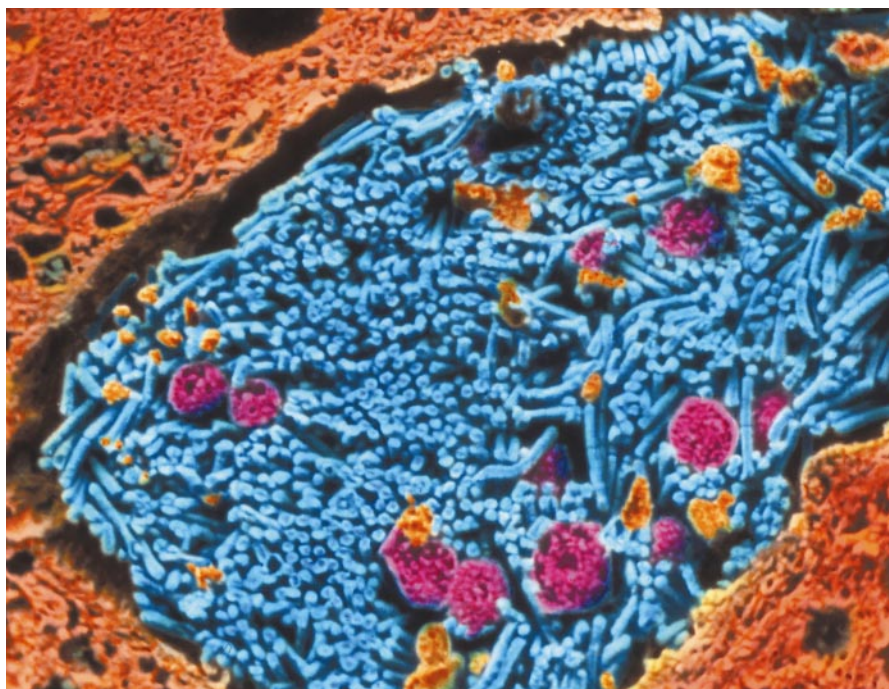
For unofficial researchers financed by grants from the MacArthur Foundation and other private sources, the team members did a staggering amount of work, and it is clear that we have them to thank for most of what we now know about the Sverdlovsk outbreak. Guillemin did the painstaking but vital 'shoe-leather epidemiology', tracking down the families of the victims and questioning them about the victims' whereabouts at the time of the aerosol release. The resulting spot map of the epidemic, showing where the deaths occurred in relation to the

suspected source of the anthrax aerosol, together with wind records covering the date of the event, pointed straight to the secret military facility, Compound 19.

When, at the end of the book, Guillemin offers her denouement, there is little doubt in the reader's mind that the 68 deaths resulted from a release of airborne anthrax spores from Compound 19 on Monday, 2 April 1979, between 1.30 and 4.30 p.m., and that the Russian physicians who presented the tainted-meat explanation were, as one of Guillemin's collaborators said, "Liars". Most of the initial mysteries are cleared up: the length of the outbreak is explained by the fact that the anthrax bacillus can remain dormant in the lung for several weeks. But other questions remain unanswered, such as why there were no child victims.

The overall set-up has all the elements of a good mystery story, but Guillemin's account does not read like one. The entire text is written in the historical present, a narrative device best used sparingly; its unrelieved use here is pretentious and makes reading the book a tedious, stultifying experience. There is a wealth of unnecessary and irrelevant detail, including accounts of some of the author's dreams (plus one of her husband's).

Scientifically, Guillemin offers two rather astonishing conclusions. First, that the victims were infected by a dosage of a mere nine spores each (the figure commonly given to



Pulmonary peril: scanning electron micrograph of anthrax bacteria (*Bacillus anthracis*) in lung.

CAMR. BARRY DOWSETT/SPL

book reviews

infect half of an exposed population is between 8,000 and 10,000 spores per individual). Second, that the total quantity of anthrax spores responsible for all 88 cases and 68 deaths, occurring over a geographical distance of approximately four kilometres (plus animal cases spread out over a much greater distance), ranges from a low estimate of "an almost unbelievable two to four milligrams, hardly enough to see", to a high estimate of slightly less than a gram (the US State Department's estimate was 10 kilograms). An alternative hypothesis, not considered by Guillemain, is that massive amounts of spores were released, infecting victims with conventional dosages, but that not everyone in the path of the cloud was exposed to the agent, or was exposed to equal quantities.

Nevertheless, this is an exhaustive, authoritative and scientifically responsible account of the largest epidemic of inhalational anthrax yet on record. ■

Ed Regis is at Western Maryland College, Westminster, Maryland 21157, USA.

An education for climatologists

Statistical Analysis in Climate Research

by Hans von Storch & Francis W. Zwiers
Cambridge University Press: 1999. 494 pp.
£65, \$110

Robert E. Livezey

An article published last year called "Statistics education in the atmospheric sciences" by Timothy Brown and four other leading meteorological and climatological statisticians lists 16 statistical books that emphasize applications to atmospheric science. Of these, only four have the breadth and internal cohesion to qualify as textbooks for classroom use. Three of these have appeared since 1994, more than 35 years after the publication of the first, the venerable *Some Applications of Statistics to Meteorology* by Hans A. Panofsky and Glenn W. Brier. The most recent on the list, *Statistical Analysis in Climate Research* by Hans von Storch and Francis W. Zwiers, is easily the most ambitious and, from the standpoint of a practising dynamic climatologist (from graduate student to senior level), the most valuable. The book is substantial in scope, rigour and its oversized format, and extraordinarily thorough in content. Very little of importance to the subject area is left untreated, which is hardly surprising considering the stature of the authors, two of the leading academicians in the science of climate variability.

The text is advanced, serving equally well as a reference, but adopts a tone and philo-

sophical stance absolutely crucial in this era of almost unlimited access to computer power, data and sophisticated analytical software. The book sets out "... to provide ... the background needed to apply statistical methodology correctly and usefully". The authors do not serve up cook-book recipes, "because they are dangerous for anyone who does not understand the basic concepts of statistics". Rather, they set out to establish the base of understanding necessary to prevent "... falling into the many pitfalls specific to our field, such as multiplicity in statistical tests, the serial dependence within samples, or the enormous size of the climate phase space". Developing respect for these difficulties for the climate scientist is the most important goal of the book.

The text is organized logically into an Introduction and six parts, entitled respectively "Fundamentals", "Confirmation and analysis", "Fitting statistical models", "Time series", "Eigen techniques" and "Other topics". Each part (except the first) is preceded by a useful overview. All but part VI (intentionally) are tightly organized, and practically constitute mini-courses, all progressing to advanced levels. The appendices are followed by a comprehensive, large reference list. The book's figures, equations and layout are attractively presented, and examples, references and cross-references are plentiful. Traditional textbook problems are not included, but examples are sufficiently frequent and well developed to compensate for this.

The book can definitely be the basis for a series of courses covering almost all of the topics suggested by Brown *et al.* for an atmospheric scientist's statistical education, lacking only material on the major topics of Bayesian inference, hierarchical Bayesian analysis, and decision theory. Among the minor omissions are cluster and discriminant analysis, whose application to climate problems has been limited recently, and, sur-

prisingly, the skewness of precipitation distributions and how they are described. Otherwise, except for topics specific to meteorological rather than climatological applications, the book is quite complete. In the context of the curriculum proposed by Brown *et al.*, this book, complemented by both Daniel Wilk's more diverse and basic *Statistical Methods in the Atmospheric Sciences* (Academic, 1995) and Edward Epstein's monograph *Statistical Inference and Prediction in Climatology: A Bayesian Approach* (American Meteorological Society, 1985), should form the centrepiece of a climate analyst's reference shelf. ■

Robert E. Livezey is in the Climate Services Division, Office of Meteorology, National Weather Service, Rm 13228, 1325 East West Highway, Silver Spring, Maryland 20910-3283, USA.

Reviving the Doctor Universalis

Albertus Magnus "On Animals": A Medieval "Summa Zoologica", Vols 1 & 2

translated & annotated by Kenneth F. Kitchell Jr & Irven Michael Resnick
Johns Hopkins University Press: 1999.
1,920 pp. \$150, £124

Cynthia M. Pyle

Albertus Magnus, who lived in 1200–80, was a prominent teacher and official of the Dominican order, and has been called the 'Doctor Universalis' in recognition of his vast learning. His writings on natural science covered physics, meteorology, geology, motion, physiology (the movement of heat and the body's 'humours'), generation, plant life and animal life. Albert the Great's treatise on animals, *De animalibus*, has now

Recent references

Companion Encyclopedia of Archaeology, Vols I & II

edited by Graeme Barker
Routledge, £160

Oxford Dictionary of Biology

edited by Elizabeth Martin & Robert S. Hine
Oxford University Press, £7.99, \$14.95 (pbk)

Encyclopedia of Geochemistry

edited by Clare P. Marshall & Rhodes W. Fairbridge
Kluwer, £280, \$480

Oxford Dictionary of Medicines

edited by Elizabeth Martin
Oxford University Press, £8.99 (pbk)

Oxford Dictionary of Physics

edited by Alan Isaacs
Oxford University Press, £7.99, \$14.95 (pbk)

Encyclopedia of the Solar System

edited by Paul R. Weissman, Lucy-Ann McFadden & Torrence V. Johnson
Academic, \$99.95, £62.95

Encyclopedia of Spectroscopy and Spectrometry, Vols I–III

edited by John C. Lindon, George E. Tranter & John L. Holmes
Academic, £570, \$874

Father of the illustrated materia medica

As a surgeon to the Roman legions of Emperor Claudius in the first century, the author Dioscorides had plenty of opportunity to travel and catalogue substances used in treating illnesses and wounds. His text *De materia medica* looks at the medical uses of more than 1,000 plant and animal products, wines and minerals. He rejected alphabetical ordering and chose to classify the material as animal, vegetable and mineral. This picture of Dioscorides comes from the facsimile edition of the superbly illustrated medieval text *Medicina Antiqua* (Harvey Miller, £48, \$75). As Peter Murray Jones points out in his introduction to this edition, Dioscorides was credited in medieval times as the innovator of illustrated *materia medica*.



been translated from the original Latin in its entirety, so that English-speaking scientists, and historians and philosophers of science, can judge for themselves the quality and substance of his scientific thought. The treatise is part commentary on Aristotle's works on the life sciences (Books I–XIX), part compendium from Thomas of Cantimpré's slightly earlier work *On the Nature of Things* (*De naturalis rerum*; XXII–XXVI), and part Albert himself (XX–XXI).

This fairly literal English translation will clearly become the standard, precisely because it is complete (Books XXII–XXVI were translated in 1987 by James Scanlan). It will allow easier access to the Latin text in Hermann Stadler's 1916–20 edition, as William Wallace notes in his rich foreword. While proofreading is sometimes lax, the translation will allow easier access to the Latin text.

The translators' Introduction reviews Albert's place in medieval science and makes a starting point for the study of his works on natural philosophy — the medieval category that included what we now call 'natural science'. The interested reader may also wish to

consult the excellent collection of articles edited by James Weisheipl (Pontifical Institute, 1980). There, and in Edward Grant's 1974 *Source Book in Medieval Science* (Harvard University Press), we can assess the extent to which Albert's work measures up to present-day standards, in fields ranging from cosmology to the life sciences.

Albert exercised his significant critical judgement of sources even when commenting on 'the philosopher', the term used in the Middle Ages to refer to Aristotle. It is important to realize, however, that although the modern reader may be struck by the breadth of what appears to be Albert's first-hand experience, this is more often a profound first- or second-hand experience of learned texts. Albert may not have done as much direct descriptive or experimental work as is sometimes claimed, even by the translators of these volumes. His reading and evaluation of the evidence, however, were careful and exceptionally critical for his time.

Albert's knowledge of hunting techniques was obtained partly from the hunters and gamekeepers he spoke to or read about (such as parts of the contemporary work on

accipiters by Frederick II Hohenstaufen), and partly from his own experience as a boy. It is highly unlikely, however, that he observed the crocodiles or ostriches he describes, since his travels (undertaken on foot, according to his vows), while extensive, were mainly in northern Europe and took him only as far south as Italy. Similarly, his detailed discussion of the anatomy of the human brain comes from Galen (whose work he knew from Avicenna's Latin translation), among other sources.

Albert's approach to the natural world shows some characteristics of a 'scientific' approach, but statements by the translators and others attributing his inability to advance beyond what he did to "historical accidents" seem like wishful thinking. A thinker must be judged in context, and Albert's powerful mind, like all minds, was subject to the intellectual norms of his time and calling. Although certainly a great and critical observer of natural diversity, he was more concerned with the central interests of natural philosophy — the overview of 'causes' and the scheme of nature as understood in the context of the medieval Latin tradition.

We can be grateful for the enormous labour expended on this full version of a difficult, historically important text by a medieval thinker of vast knowledge and unquestionable intellectual power and dedication.

Cynthia M. Pyle is at 470 West End Avenue, New York, New York 10024, USA.

Something for everyone

An Illustrated Guide to Theoretical Ecology

by Ted J. Case

Oxford University Press: 1999. 464 pp.

£22.50, \$49.95

Joan Roughgarden

Thirty years ago, the phrase theoretical ecology seemed a contradiction in terms. Today, though, theoretical ecology has matured, and is now represented by a curriculum shared by most courses in the subject at universities throughout the world.

Ted Case, a distinguished theoretical ecologist and professor of biology at the University of California at San Diego, has written the most recent textbook to serve this market. It is by far the most attractive book yet, and readily earns its title of being an 'Illustrated Guide'.

But the book is not only replete with pictures. It offers rigorous derivations of the models and theorems in theoretical ecol-

book reviews

ogy, using all the usual mathematics from calculus, linear algebra and differential equations. But it stands out because of the completeness of its graphical illustrations, which are so extensive that a reader may follow the derivations solely from the graphs with little reference to the mathematical formulae.

The topics covered are those found in most courses on theoretical ecology: density-independent and density-dependent population dynamics, age-structured population dynamics, life-history theory, species interactions, multi-species communities, and island biogeography. Spatial population processes are nicely worked in throughout all the chapters. Practical data, too, are frequently mentioned, and used for illustrative calculations.

Students in theoretical ecology typically come to the subject with a diversity of mathematical backgrounds. Some may have had only one, dimly remembered, course in college calculus. Others may be math whizzes in linear algebra, differential equations or computer programming. All will find this book helpful and relevant. The material is divided into normal and advanced sections, so both the undergraduate and graduate student will find it of value.

Instructors may wish to supplement this book with more specialized material on optimal foraging theory, kin selection, game theory, the trophic cascade, symbiosis/mutualism interactions, landscape and ecosystem ecology, and Earth-systems models of coupled physical/biological processes. I personally also like the Gaia models, which, though speculative, emphasize the importance of the biosphere in planetary dynamics. This is not to suggest that these topics should have been included in the book. It already has 449 full-size pages, and could not be much bigger without becoming unwieldy.

All in all, the *Illustrated Guide* is an accessible and thorough text. It is a welcome addition to the book list for theoretical ecology, and I recommend it to advanced undergraduate students, graduate students and professionals in ecology, conservation biology and resource management. ■

Joan Roughgarden is in the Department of Biological Sciences, Stanford University, Stanford, California 94305, USA.

More on ecology and the environment

Dynamics of Coral Communities

by Ronald H. Karlson

Kluwer Academic, £88, \$150

The Effects of UV Radiation in the Marine Environment

edited by Stephen de Mora, Serge Demers & Maria Vernet

Cambridge University Press, £50, \$80

Science in culture

Lancing lasers

Nam June Paik at the Guggenheim Museum.

Martin Kemp

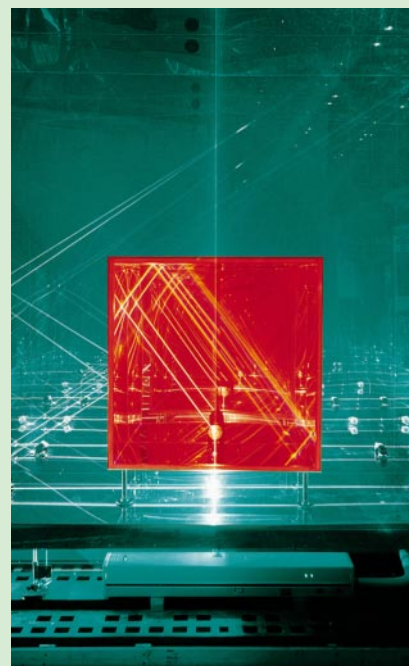
Since his explosion into the anarchic avant-garde of the Fluxus movement in Germany and New York during the early 1960s, no *enfant* has been more inventively *terrible* than the Korean master of electronic and performance art, Naim June Paik (born in 1932). Even his severe stroke in 1996 has not diminished his creative energies and desire to push new media to their technological limits. For his current show at the Solomon R. Guggenheim Museum in New York (until 26 April), he has activated the dynamically challenging space of Frank Lloyd Wright's helical gallery with spectacular laser installations, alongside his more familiar compositions of multiple television monitors.

Strongly associated with the leading musical experimentalists of the late 1950s, Karlheinz Stockhausen and John Cage, Paik became centrally involved in a series of iconoclastic performances that assaulted audiences' comfortable expectations about music, art, dance and theatre. Pianos and violins were shockingly smashed as spectators winced. In 1971 the cellist Charlotte Moorman performed provocatively topless, apart from Paik's brassiere of two small TV screens. It was in the more technological dimensions embodied in the bra that Paik's future was to lie.

By 1963 Paik had already decided to dedicate his prodigious talents to "the spartan life of physics and electronics". The cathode-ray tube, which was becoming unthinkingly naturalized in popular culture, was chosen as his primary medium. By turns he subverted the TV set, satirizing its dehumanizing technology, and exulted in its potential as the true medium for the late-twentieth-century creator. Why, he asked, should the artist not be able to "shape the TV screen canvas, as precisely as Leonardo, as freely as Picasso, as colourfully as Renoir, as profoundly as Mondrian, as violently as Jackson Pollock, as lyrically as Jasper Johns".

In collaboration with the Japanese electronics pioneer Shuya Abe, he developed increasingly sophisticated versions of the 'Paik Abe video synthesizer', which allowed video images to be manipulated, formed into collages and simultaneously displayed on multiple monitors, or composed symphonically into restless kaleidoscopes of electronic imagery, both abstract and figurative. Alongside such images transmitted through TV sets, he also worked directly on the cathode-ray tube itself, using magnets to contort the configurations of rays on the screens into strange geometries, akin to the wave motions of an oscilloscope.

The laser-works that he has recently undertaken in collaboration with Norman Ballard, most notably the triptych of the *Three Elements*—in the form of a triangle, circle and



Detail of *Three Elements* by Naim June Paik with Norman Ballard—lasers, mirrored chambers, prisms, motors and smoke.

square—are a natural continuation of such compositions as his *Magnet TV* of 1965. Within the shallow, mirrored cabinets of *Three Elements*, triangular and square prisms rotating in different phases stab thin laser beams into an ether of pale vapour. Penetrating tracers of arrowed light criss-cross the elemental figures. Scintillating points oscillate along defined tracks at the edges of each shape and at the intersections of the dancing beams. The razor-sharp geometries, refracted through the prisms and reflected off the containing mirrors at ever-changing angles in endless variations, are etched across an apparently unconfined space, as a result of internal reflections off the two-way (50%) mirror at the front of each cabinet.

In the modernist museum to cap all modernist museums, the basic geometries pay open homage to the canonical simplicities of minimalist abstraction, not least Josef Albers's series on "Homage to the Square" (see *Nature* 390, 451; 1997). But, as might be expected of an artist of such global experience and concerns, the resonances are much more extensive than an incestuous reference to earlier art. As alert to Platonic idealism as is Zen Buddhist meditation to eternal simplicities, Paik uses a technology perfected by late-twentieth-century physics to transform ancient archetypes into a vision of mathematical energies transmitted across infinite space. It is a vision worthy of our entry into the new millennium. ■

Martin Kemp is in the Department of the History of Art, University of Oxford, 59 George Street, Oxford OX1 2BE, UK.

Abstraction and idealism

From Plato to Einstein: how do we acquire knowledge?

Semir Zeki

Brain studies are poised to occupy a dominant scientific position during this century, but one of the enduring problems was defined over two millennia ago by Plato and reinforced by his successors in the Western philosophical tradition. The problem revolves around the acquisition of knowledge, a primordial function of the brain. Impressed by the Heraclitan doctrine of flux or constant change, Plato and his successors sought to understand how we can obtain knowledge about the permanent and non-changing properties of all that is around us, when the information reaching us is changing all the time and when we ourselves are subject to change. With the exception of Schopenhauer, they did not cast the problem in the context of the complex organization of the sensory brain, because they imagined that 'real' things (Platonic ideals and the Kantian 'thing-in-itself') belonged to a supra-sensible world accessible through thought alone. Kant sought to define the limitations imposed by the mind in the acquisition of knowledge, and the formal contribution that it makes to that acquisition. He believed that there are two innate intuitions, time and space, into which all experience is read. In more recent times, Einstein emphasized the importance of the thought process in acquiring knowledge when he wrote that "the critical thinking of the physicist cannot ... proceed without considering critically a much more difficult problem, the problem of analysing the nature of everyday thinking".

Implicit in these writings is the supposition that a fundamentally similar thought process

underlies the acquisition of all knowledge. Yet neurology has taught us that knowledge, in terms of both its acquisition and the processes underlying it, is modular and distributed. Even within the visual system, knowledge about the colour of surfaces, for example, is acquired by a different neural process from the process dealing with visual motion. This can have bizarre consequences, as when lesions occurring in the colour centre lead only to an incapacity to acquire knowledge about colour, leaving the sensing of motion, for example, unimpaired. Moreover, the 'innate intuitions' of time and space for constructing colour and motion, although they share similarities, are also significantly different: colour requires the brain to determine the amount of light of different wavelength reflected from different surfaces simultaneously, whereas successive stimulation in time is essential to the motion system.

Where, then, can we seek that unity of thought processes which constitute the foundations of all knowledge? I believe that there are two closely linked and automatic processes — abstraction and the formulation of ideals — which underlie our ability to acquire all knowledge because they are the characteristic features of any efficient knowledge-acquiring system. The former is both selective and eliminative; it allows the brain to determine some property or relation which is common to many particulars, thus making it independent of the particular. Abstraction is also imposed on the brain by the limitations of its memory system, since it does away with the need to recall every detail. Memory is a critical step in obtaining knowledge, but, as Descartes saw,

it could not be trusted in an unqualified way, even within the certain

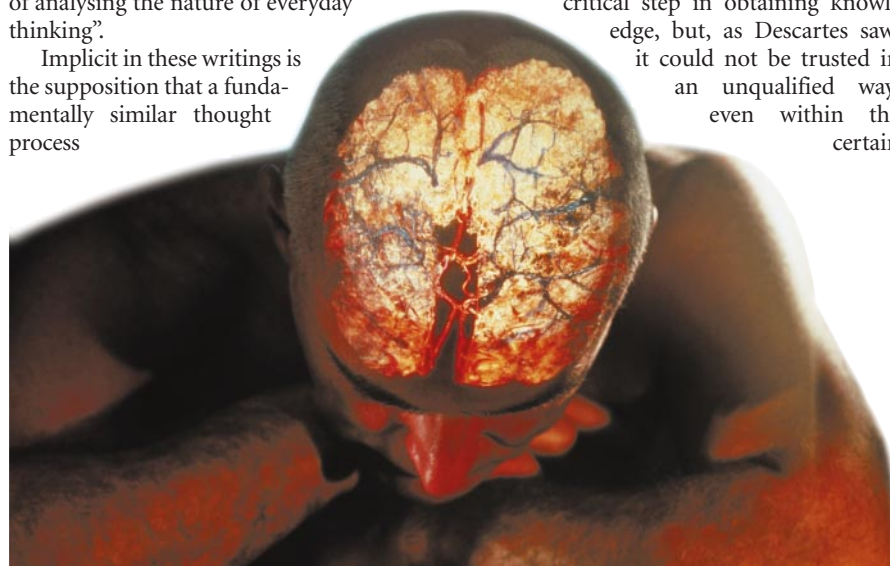
Kant believed that there are two innate intuitions, time and space, into which all experience is read.

world of mathematics. A cardinal step in mathematical formulations is the deductive one. But in deductive arguments, the mind has to rely on memory in tracing the earlier steps in a deductive process. And the memory process itself can be at fault (as it often is).

Abstraction leads naturally to the formation of ideals. Plato used the term 'ideal' to mean a universal — derived from the intellect alone — as opposed to the particular, derived from sensory experience. Because memory of the particular fades, the ideal built by the brain from many particulars becomes the only real thing about which we can have knowledge, much as Plato and Kant believed. I would depart from their ideas by saying that the 'thought processes' involved in generating brain ideals are automatic neural processes, which differ according to the ideal that is being constructed, but that each of the many distributed knowledge-acquiring systems of the brain has an independent capacity to form ideals. We are not conscious of the automatic neural thought processes that are necessary for us to perceive the reality behind all natural phenomena, although the results of the processes are consciously perceived. This leads to my supposition, modified from Leibniz, that a different unconscious process underlies each of the conscious events that we experience.

A study of the neurological basis of abstraction and ideal formation is now within our reach. It may also shed light on another issue, namely the psychological 'suffering' which Freud traced to our 'mental constitution'. If, as I believe, abstraction and the formation of ideals are the necessary characteristics of efficient knowledge-acquiring systems, and if knowledge can be acquired only through particulars — which is the staple diet of the brain — then it is not hard to see that there will forever be a clash between the ideals built from many particulars and the experience of the particular, leading to a perpetual dissatisfaction. ■

Semir Zeki is in the Wellcome Department of Cognitive Neurology, University College London, Gower Street, London WC1E 6BT, UK.



Abstract thinker: the intellectual ideal clashes with sensory experience.

The new laureate speaks

The response of this year's Nobelista in full

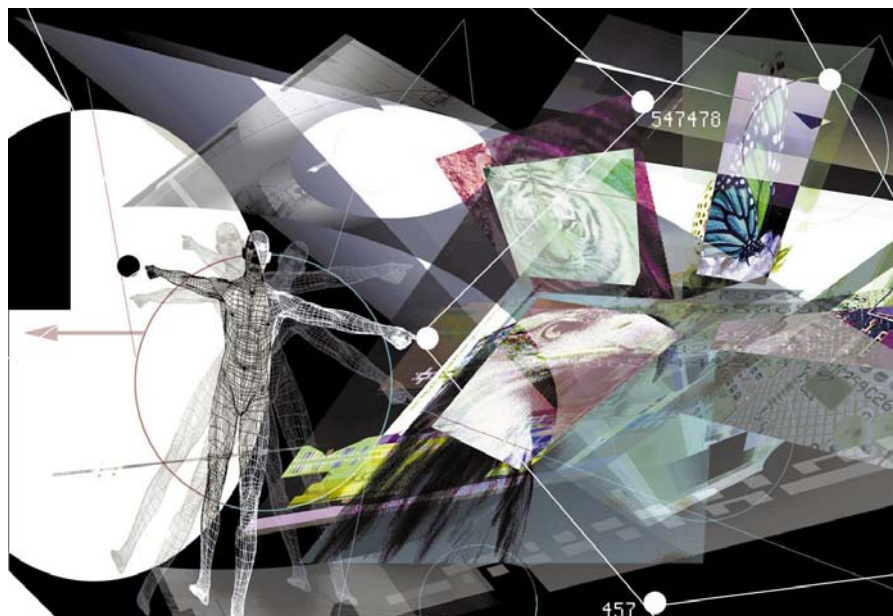
Charles Sheffield

It is customary, and almost mandatory, in accepting awards such as this to say that one's own efforts rest on the backs of the giants of earlier days. I would prefer, with all modesty, to say that my work and the work of my colleagues has depended on spinners; the spinners, in particular, of two great threads of scientific history.

My first thread concerns the mutability of species. For most of human history, the forms that inhabit Earth's lands and seas were regarded as unique, well-defined, and unchanging. The alternative view, first advanced over a thousand years ago by Charles Darwin and Alfred Russel Wallace, produced intellectual shock waves that have affected science and society everywhere (with the possible exception of Kansas, which holds out to this day.) Species can and do change, through natural selection, to produce new species. The timescale for such events is long, measured in millions of years, in contrast to the 'unnatural selection' that the farmers and breeders of the planet have used for ten or so millennia to produce preferred varieties of plants and animals. Evolution, however, is inevitable and continues to this day.

A second shock came less than a century later, when in 1953 Crick and Watson elucidated the molecular structure of DNA. As its central role in life became clear, so did human ability to map and modify its structure. Through the wholesale mix-and-match of genetic materials, almost any combination of physical characteristics became possible. Moreover, the timescale for development of these hybrids could be measured in years or months, not aeons. In consequence, today the idea of a biological 'species' is still a useful concept, but the present definition of the word would baffle Darwin's predecessors. In their day, the biological species resembled integers along the real number line, well defined and well separated. Now, between any pair of those original species, there exists a whole continuum of intermediate species created routinely by DNA manipulation. The single points have become a continuous line. I call this the life-line, upon which my own work has heavily depended.

The second thread derives from exactly the same period as Crick and Watson's work. Despite Charles Babbage's attempts to construct a mechanical analytical engine more than a century and a half earlier, the first



successful computer awaited the development of reliable electronics. From humble beginnings in the 1940s, it rapidly became an astonishing success story that went from exotic rarity to ubiquitous presence in less than a human lifetime of the day.

We can liken the earliest computers, such as the ENIAC and EDSAC, to primitive life forms. Successive generations evolved fast and far in both hardware and software, so that within fifty years the sleek laptops and supercooled teraflop machines seemed as remote from the original ur-computers as an amoeba is different from an elephant. However, the amoeba and the elephant share strong family resemblances at the DNA level. At a similar deep level the logic was the same in all early computers, and the lineage of those first machines was clear (when our historians plumb the depths of Windows 98, they find within it the primitive reptilian brain of DOS).

The early computers offered an apparent choice among hundreds of models, but most of the variability was superficial. Discount the differences in speed, memory, and operating systems, and the enumeration of the distinct computer 'species' at the turn of that millennium needed no second digit in decimal notation. Like the biological species, those early computer designs formed isolated points along a line of their own. I term this the machine-line. And today our computers, unrecognizably different in size, speed, and capability from their distant ancestors, populate the full machine-line continuum of computational types and capabilities.

Now it is time to weave the two threads together. This interweaving of life-line and machine-line at first came slowly. The initial steps were primitive, in spoken inputs to computers, grammar and spelling checkers, on-line patient care clinics, and 'smart' pace-makers. The next steps followed within a couple of human generations: eye lenses changing focus as needed, or boosting contrast to an aging retina. They contained the so-called 'supercomputers' of their day, which it is all too easy for us to mock. There were similar hearing aids, with full directional sensitivity and selected frequency enhancement; and, of course, there was computer activity directed by built-in brain implants, permitting communication with people mentally 'locked in' through accident or disease, and in many ways the direct precursors to my own work.

In truth, what I and my colleagues did was simple. Indeed, to anyone with mathematical tendencies it is almost inevitable. Let us retain the life-line of the biological species as the real coordinate axis, and let us assign the machine-line of computer types to an imaginary coordinate axis. We now have a 'complex plane of organisms', any point of which represents some combination of organic and inorganic organized systems. Call each point a cyborg. It is for the creation of the 'theory of functions of a complex variable', appropriate to the description and analysis of all possible cyborgs, that I and my human companions feel honoured to be rewarded today.

Charles Sheffield's latest book is *Borderlands of Science*. He also writes a weekly column.

When Greenland ice melts

Christine Schøtt Hvidberg

Sea levels are likely to rise in response to global warming. But by how much? Identification of the causes of the six-metre surge during the last interglacial provides a useful clue.

Glaciers and ice sheets grow and shrink in step with climate changes. In turn, global sea level varies according to the amount of water stored as ice in glaciers and ice sheets or released as melt water. For example, during the last interglacial (130,000 to 110,000 years ago) the climate was warmer than today, and the global sea level was around six metres higher than at present. Then, during the last ice age, enormous ice sheets grew in the Northern Hemisphere, and at their maximum (about 20,000 years ago) the sea level was around 120 metres lower than it is today. A question being considered in the debate over global climate change is the threat of a rapid increase in sea level owing to the possible collapse of the West Antarctic Ice Sheet. Complete melting of the West Antarctic Ice Sheet today would lead to a sea level rise of around six to seven metres accompanied by catastrophic global flooding. Because of the high sea level during the last interglacial, it was suggested that the West Antarctic Ice Sheet may have disintegrated completely, and could do so again with global warming.

Currently three-quarters of the Earth's fresh water is locked in the great ice sheets covering Greenland and Antarctica. On the basis of new knowledge of the climate history in Greenland from ice cores, Cuffey and Marshall¹ (page 591 of this issue) show that the Greenland Ice Sheet was much smaller during the last interglacial than previously thought, and contributed at least four or five metres to the peak sea level rise (Fig. 1). Their result implies that high sea levels during the last interglacial should not be interpreted as evidence for extensive melting of the West Antarctic Ice Sheet, and so challenges the hypothesis that the West Antarctic is particularly sensitive to climate change.

The West Antarctic Ice Sheet is partly a floating ice shelf surrounded by rocky islands and partly an ice sheet grounded below sea level. It is the melting of the overland ice that contributes to rising sea levels — floating ice already displaces its weight in sea water, so its break up won't directly affect sea levels. The flow of ice from the inland part of the West Antarctic Ice Sheet to the ocean is concentrated in fast-moving glaciers (ice streams). The interplay between the ice streams, the ice sheet geometry and nonlinear ice flow is not yet fully understood, making the outflow of ice extremely difficult to



Figure 1 A small fishing boat is dwarfed by icebergs in the Jakobshavn ice fjord, West Greenland. Despite the apparent stability of the Greenland ice sheet, Cuffey and Marshall¹ show that the melting of Greenland ice made a substantial contribution to high sea levels during the last interglacial.

model. The West Antarctic Ice Sheet formed several million years ago, but it is unclear how much its size has fluctuated since then. In particular, its stability during the last interglacial has generated much interest. If it collapsed then — a period with global temperatures of maybe only 2 °C warmer than today — it could disintegrate again as a result of global warming. But its stability in response to climate changes remains unclear².

The Greenland Ice Sheet contains enough ice to raise the global sea level by six to seven metres, about the same as the West Antarctic Ice Sheet. But unlike the West Antarctic it was always thought to be insensitive to climate change. This is because the ice sheet is grounded above sea level, and is expected to respond gradually to climate change by melting at its surface. Previous models of the Greenland Ice Sheet indicated that it contributed around one to two metres to the sea level rise during the last interglacial, and that the surface height of central Greenland has been almost unchanged during the past 150,000 years^{3,4}.

Predicting how an ice sheet evolves over time requires an assumed climate history, which is fed into the numerical model. The Greenland climate history is derived from

the record of oxygen-isotope ratios from central Greenland ice cores⁵, which are then converted to temperatures. This conversion is not trivial. The oxygen-isotope ratio of water falling as snow on the ice sheet depends not only on surface temperature, but also on atmospheric transport of the vapour, the surface height and a number of other factors. Until recently, oxygen-isotope ratios were converted to temperatures by assuming the relationship that exists today between oxygen-isotope ratio and surface temperature at different sites on the Greenland Ice Sheet.

Temperature measurements from central Greenland bore holes have recently provided a direct and independent measure of Greenland's temperature history^{6–8}, and thereby a calibration of the oxygen-isotope record back through some 100,000 years. Temperatures older than that cannot be detected in the Greenland Ice Sheet. The results reveal that the difference between the temperature minimum during the last ice age and today was 20–25 °C in central Greenland, in contrast to the previously assumed 10 °C, implying that oxygen-isotope ratios are less sensitive to temperature changes than was thought. Cuffey and Marshall's new picture¹ of the Greenland Ice Sheet, which suggests extensive melting during the last interglacial,

BRYAN AND CHERRY ALEXANDER PHOTOGRAPHY

is primarily a result of using the new recalibration of the oxygen-isotope climate record.

Cuffey and Marshall make the crucial assumption that the new oxygen-isotope temperature calibration holds back into the last interglacial, beyond the period covered by the bore-hole temperature records. The bore-hole record covers the present interglacial and part of the last ice age, that is, two periods with widely different climates and sea levels. So far there is no reason to believe that this relationship will not hold for older ice. The implication is that the last interglacial was 5–10 °C warmer in central Greenland, instead of 4–5 °C as previously thought.

Modelling the Greenland Ice Sheet is further complicated by uncertainty surrounding the stability of the climate in Greenland during the last interglacial⁵. For temperatures older than 98,000 years, Cuffey and Marshall use a constructed record based on the oxygen-isotope record from an ice core from Vostok in Antarctica. A climate record cannot simply be transferred from one location to another. But a number of sensitivity experiments confirm that the ice sheet retreat occurred within a few millennia in the warm interglacial climate (a period shorter than the expected duration of the last interglacial). Therefore the result is not very sensitive to the duration and course of the assumed climate history, and it is safe to use the record from Vostok.

Understanding the causes of the high sea

level during the last interglacial, which actually represents a warmer climate than today, is clearly relevant to debates about the consequences of global warming. For coastal populations living with the threat of flooding, it doesn't matter whether the melt water comes from Greenland or Antarctica, but our ability to predict how fast these ice sheets will shrink is crucial. Cuffey and Marshall estimate that the widespread melting of the Greenland Ice Sheet took place within a few millennia, but they draw a less dramatic picture than the suggested collapse of the West Antarctic Ice Sheet, with its accompanying rapid increase in sea level. Perhaps the threat of such a sudden collapse is exaggerated. The work of Cuffey and Marshall is based on our limited knowledge of climate history. Continuing deep drilling in North Greenland, the North Greenland Ice Core Project⁹, might help to shed further light on climate history and possible patterns of future climate change.

Christine Schøtt Hvidberg is in the Department of Geophysics, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark. e-mail: ch@gfy.ku.dk

1. Cuffey, K. M. & Marshall, S. J. *Nature* **404**, 591–594 (2000).
2. Oppenheimer, M. *Nature* **393**, 325–332 (1998).
3. Letréguilly, A. *et al. Palaeogeog. Palaeoclimat. Palaeoecol.* **90**, 385–394 (1991).
4. Fabre, A. *et al. Ann. Glaciol.* **21**, 1–7 (1995).
5. Grootes, P. M. *et al. Nature* **366**, 552–554 (1993).
6. Cuffey, K. M. *et al. Science* **270**, 455–458 (1995).
7. Dahl-Jensen, D. *et al. Science* **282**, 268–271 (1998).
8. Johnsen, S. J. *et al. Tellus* **47B**, 624–629 (1995).
9. Dahl-Jensen, D. *et al. J. Glaciol.* **43**, 300–306 (1997).

Neurobiology

The good taste of genomics

Stuart Firestein

Be it an expensive dinner or a quick snack, the complexity of a flavour results from the multimodal nature of the perception. A flavour is affected by smell, touch and even pain (hot and cold) inputs, all of which are wedded to the tastes detected by sensory cells on the tongue. The four primary tastes are salty, sour, bitter and sweet. More recently, a fifth taste — umami — has come to be recognized as an independent sensation; this is common in Asian cuisines and is the taste of the amino acid glutamate. But how are these tastes detected? Thanks to a remarkable concurrence of papers^{1–3}, we are now several steps closer to answering this question. Matsunami and colleagues, writing on page 601 of this issue¹, and Adler *et al.* and Chandrashekar *et al.*, reporting in a recent issue of *Cell*^{2,3}, have been mining genomic data, and have struck gold with their discovery of a new family of bitter-taste receptors in mice and humans.

The biochemical and physiological

mechanisms by which tastes are detected and discriminated are quite different. Salty and sour tastes, for example, result simply from the influx of sodium ions (for salty tastes) or hydrogen ions (for sour tastes) through channels in the uppermost membranes of taste-receptor cells (TRCs). By contrast, the sweet, bitter and umami tastes were thought to be detected by receptors that bind to a specific taste molecule. These receptors were presumed to be linked to guanine-nucleotide-binding (G) proteins, well-known components of intracellular signalling cascades. However, the precise mechanisms involved and the identities of the molecular players remained unclear. This is where the new papers^{1–3} come in. As predicted, the bitter-taste receptors in mice and humans do indeed make up families of previously unknown G-protein-coupled receptors (GPCRs). Why has this result been so long in coming, and what was the breakthrough?

The problem has been that most of the tongue has nothing to do with taste. TRCs are packed into small, onion-shaped structures called taste buds, with from 50 to 100 cells in a single bud. There are only a few thousand taste buds and perhaps 30,000–50,000 TRCs, which are hard to isolate and do not provide much material with which to work. The solution has been provided by the progressive sequencing of the human and mouse genomes, and the ever-increasing amount of sequence data available in public databases.

Starting in humans, Matsunami *et al.*¹, Adler *et al.*² and Chandrashekar *et al.*³ had an important lead: there is a known genetic variation that results in reduced sensitivity to a bitter substance known as PROP. This variation maps to a particular region of the human genome for which a significant amount of sequence was already available. This is where the hunt for genes encoding taste receptors began. The hypothesis was that, because there are a lot of chemically different bitter substances (Fig. 1), there would be a large family of receptors to recognize them; this family might be encoded by a gene cluster, mapping to a tight region of the genome, perhaps near the known site of variation. Indeed, the authors did identify candidate genes for taste receptors in this region (in retrospect, this was probably fortuitous because the PROP-sensitivity gene turns out not to encode a taste receptor). Using the sequences of these genes to search further in the genome, the groups identified another 19 related genes, spread over three chromosomes. A new family of receptors was born.

A cluster of genes encoding GPCRs also mapped to a region of the mouse genome homologous to that of the human region. This allowed the critical molecular proof to be developed in mice. The messenger RNA encoding the receptors was present in a subset of TRCs in mice, and expression of these receptors was highest in taste tissue. Finally, when the receptors were expressed in cultured mouse cells, the cells responded to specific bitter substances. This strategy makes plain the importance of having multiple genome sequences available for research. Rigorous proof of the function of these genes would have been nearly impossible to obtain in humans.

But have all the bitter-taste receptors now been identified? The answer is probably not, as only 14% of the human genome was available to be searched. Taking into account the existence of 'pseudogenes', and the fact that the taste-receptor genes are probably not distributed randomly across the genome, there could be between 50 and 90 receptors in all.

Interestingly, it seems that a given taste cell makes many, perhaps all, of the receptors. If a single cell makes many receptors, it

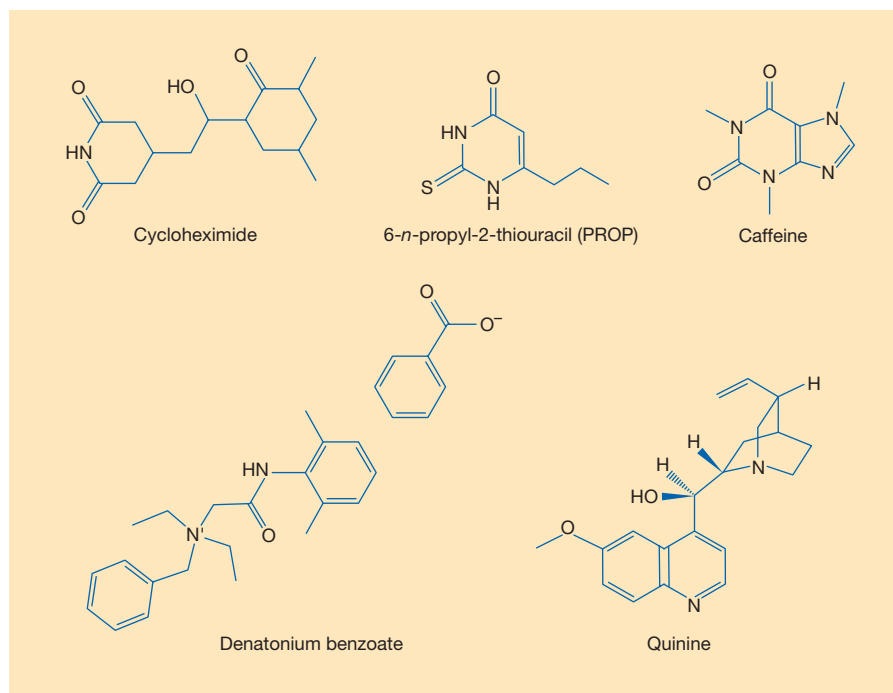


Figure 1 Leaving a bitter taste in the mouth. The chemical structures of even these few bitter substances are very diverse. This leads one to suspect that there must be a family of bitter-taste receptors in the tongue — a family that has now been identified by Matsunami *et al.*¹, Adler *et al.*² and Chandrashekar *et al.*³.

will be sensitive to many bitter compounds, but the brain will not be able to distinguish which particular bitter substance is being detected. Indeed, humans often cannot distinguish between bitter stimuli. Bitter substances are frequently poisonous or harmful and need to be detected at low concentrations, before a fatal amount of something is swallowed. But because bitter substances include many different types of chemical (Fig. 1), a single receptor that recognized many bitter compounds could not be very sensitive to any one of them. The strategy of expressing multiple receptors, each recognizing a specific bitter taste, results in TRCs with a broad range but high sensitivity.

Victor Hugo once claimed that nothing is as powerful as an idea whose time has come, and this certainly seems to be true of taste receptors. Just a few weeks ago came the report⁴ of a new family of receptors in the fruitfly *Drosophila melanogaster* that were distinct from the fly odour receptors but were also found in chemosensory organs. These new receptors seem to be the fly taste receptors, another distinct family of GPCRs. Finally, there is the curious taste umami — the taste of glutamate. We have glutamate receptors in the brain, for glutamate is an important neurotransmitter. But glutamate activates the brain glutamate receptors, called mGluRs, at concentrations that are far below taste thresholds. Chaudhari *et al.*⁵ reported recently that an alternatively spliced isoform of mGluR4, with a more appropriate affinity for glutamate, is

expressed in a subset of TRCs, so this may be the umami detector.

The taste field has had quite a start to the millennium. Where do we go from here? Taste is important in many areas of our lives. Caloric intake and salt ingestion are two obvious areas that might benefit from an understanding of the underlying taste mechanisms. Many medicines have a terribly bitter taste — so much so that patient compliance is often compromised. The development of bitter antagonists is now within reach, allowing us to increase the palatability of medicines and even of foods that are 'good' for you. Insect pests identify feeding and breeding plant hosts by taste, raising the possibility of controlling such pests by tastes rather than by toxic and environmentally harmful insecticides. Of course these developments are not trivial, and nor are the remaining questions — the identity of the sweet receptor, for example. But with the receptors that we now have and the application of a similar strategy to the newly available genome data, it should be well within our power to lick these problems. ■

Stuart Firestein is in the Department of Biological Sciences, Columbia University, 923 Fairchild, MC 2438, New York, New York 10027, USA. e-mail: sjf24@columbia.edu

1. Matsunami, H., Montmayeur, J.-P. & Buck, L. B. *Nature* **404**, 601–604 (2000).
2. Adler, E. *et al.* *Cell* **100**, 693–702 (2000).
3. Chandrashekar, J. *et al.* *Cell* **100**, 703–711 (2000).
4. Clyne, P. J., Warr, C. G. & Carlson, J. R. *Science* **287**, 1830–1834 (2000).
5. Chaudhari, N., Landin, A. M. & Roper, S. D. *Nature Neurosci.* **3**, 113–119 (2000).



100 YEARS AGO

The question as to the origin of the energy possessed by the Becquerel rays is one of considerable interest. The existence of substances capable of emitting radiations possessing energy, without any appreciable loss of weight or introduction of work from external sources, would appear to be impossible from the view of conservation of energy. The measurements of M. Henri Becquerel upon the deviation of the radium rays in an electric field, taken in conjunction with those of M. and Mme. Curie of the charges carried by these rays, lead to results which show a way out of the difficulty, on account of the extreme minuteness of the quantities of energy in question. The calculations of M. Becquerel show that this energy is of the order of one ten-millionth of a watt per second. Hence a loss of weight of about a milligram in a thousand million years would suffice to account for the observed effects, assuming the energy of the radium to be derived from an actual loss of material.

From *Nature* 5 April 1900.

50 YEARS AGO

Preliminary trials during the past two years, which aimed at replacing costly cultivations in root crops by chemical methods of weed control, have now been completed at Jealott's Hill. *iso*-Propylphenylcarbamate, 'Methoxone' and 2,4-dichlorophenoxyacetic acid were known to prevent the germination of certain plant species when applied as pre-emergence dressings, and the species response to *iso*-propylphenylcarbamate was different from that of the other two compounds. Therefore, mixtures of *iso*-propylphenylcarbamate with 'Methoxone' and 2,4-dichlorophenoxyacetic acid were examined alongside pre-emergence dressings of the individual components. The applications were made at different intervals prior to sowing, and these were followed by observations on subsequent crop-growth and weed-control. The main experiments were on kale and mangolds; but lettuce, onions, field beans, peas, lucerne, sugar beet and swedes were included in the second year. The preliminary trials are being extended to cover a range of soil and climatic conditions; but, although the conclusions are only tentative, they point to new possibilities of weed control in crops where at present few chemical methods are available.

From *Nature* 8 April 1950.

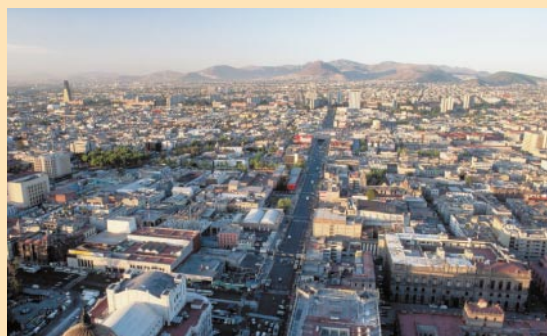
Urban climate

The water cooler

Ever thought about escaping the heat of the city by getting down to the lakeside?

According to Arón Jazcilevich and colleagues (*Climatic Change* **44**, 515–536; 2000), having a lake on your doorstep may not only provide a waterside retreat but help keep the city's climate appreciably cooler.

Mexico City, it seems, was such an example. Earlier this century, the relatively small city (86 km²) was bordered by Lake Texcoco (120 km²), and the 'urban heat island' effect — where city temperatures exceed those in surrounding rural areas — was only 1.5 °C. Urban expansion has dramatically altered the balance, so that the sprawling city (pictured), at 1,200 km², now dominates the Basin of Mexico. The lake has been reduced to a mere 10 km² or so in area, and the heat of the city can exceed that in neighbouring countryside by



8–10 °C. This commonly observed effect is usually explained by the surface geometry and thermal properties of all that extra concrete and asphalt.

But what part might the lake play? Jazcilevich *et al.* developed a model to reproduce urban climate under present-day conditions and those of 1921, using the respective city and lake areas. Then, taking the situation in 1921, and reducing lake area to today's extent, they found that the expected temperature

field was almost identical to that of today. From this the authors infer that the increase in city temperatures is largely due to the reduction in lake area and the evaporative cooling it provided, as well as the massive urbanization.

Perhaps the finding has come too late for Mexico City. As urban areas continue to swallow up land around the world, however, town planners would be wise to keep a little bit of Venice in their designs for the comfort and health of inhabitants.

Jim Gillon

Evolution

Bacterial cheaters

Joan E. Strassmann

Cooperative groups of higher organisms are vulnerable to cheaters that reap the benefits of cooperation without paying the costs. For example, an unrelated male lion may take a gazelle away from the females that killed it. On page 598 of this issue¹, Velicer and colleagues show that similarly selfish behaviour is also seen in bacteria — specifically, in groups of *Myxococcus xanthus*.

As they move through the soil, preying upon other microorganisms, *M. xanthus* individuals behave in group-coordinated ways that are more usually associated with complex social animals than with bacteria^{2,3}. Individuals move and feed as a cooperative, tight group reminiscent of teams of hunting army ants or wolves, but in this case secreting enzymes that kill and degrade (lyse) the prey before them^{2–4}.

But *M. xanthus* display their most dramatic social act when they are starving. Individuals aggregate even more densely than

when feeding and form a raised 'fruiting body' (Fig. 1, overleaf), in which a minority of cells convert from vegetative rods to hardy, spherical spores^{2,3}. During this process, many cells commit suicide and lyse, releasing their contents⁴. The contents of lysed cells might enable spore formation by other cells — the contents may be eaten, or might otherwise influence the sporulating cells⁴. The development of a fruiting body involves cell-to-cell signalling, the formation of a multicellular structure, and morphological changes in individual cells. Myxobacteriologists therefore consider the fruiting body to be a simple model system in developmental biology³, with processes at least partly analogous to those that form structures as complex and integrated as a fruitfly's body.

Standing in marked contrast to this cooperative, and often altruistic, behaviour, however, are the social cheaters described by Velicer *et al.*¹. The authors isolated *M. xanthus* mutants showing developmental

defects from six populations that had evolved in a liquid medium in which fruiting-body formation could not occur. The populations had evolved for 1,000 generations — a short time in evolutionary terms, but still a much longer period than would be feasible in experiments on most organisms. After this period, clones from all six mutant lines showed a decreased ability to form spores when starved on a solid substrate, a loss of function that is not surprising given that the bacteria had adapted to an environment in which the function is not needed.

Velicer *et al.* then mixed the mutants with ancestral (wild-type) populations, at a ratio of 1% evolved clone to 99% ancestral clone. Under these competitive conditions, five of the six defective clones produced more spores, relative to the number of input cells, than they had when in isolation. Three of the clones did even better in such mixtures than the wild-type clones, and thus qualified as 'cheaters'.

Velicer and colleagues also studied three other, previously known, developmental mutants that were defective in either of two signalling systems and, consequently, in spore production when in pure culture. In mixtures of 1% mutant to 99% wild-type clones, two of these mutant clones generated spores at a higher frequency than did the wild-type clones. So, some developmental mutants not only revert to spore production when in a mixture with wild-type cells, but can even be over-represented in the spores in comparison with their original frequency.

But exactly how these mutants cheat is not obvious. There is no clear functional class of altruists, analogous to honeybee workers or to stalk cells in cellular slime moulds, that the cheating *M. xanthus* avoid joining. The *M. xanthus* individuals that are the best candidates for altruists are the cells that lyse during development, although it is not even certain that these lysed cells are actually eaten by the sporulating cells. Moreover, we do not know whether the cheating clones cheat specifically by avoiding lysis. The mutants might simply be able to continue growing while some of the wild-type cells lyse, because the mutants do not pay all the metabolic costs of producing the signalling chemicals that contribute to the formation of the fruiting body⁵. If the mutants' advantage is indeed that they can simply keep on growing, they would need to be able to continue dividing after being plated on starvation medium. This might be possible, as they could eat lysed cells.

Cheating *M. xanthus* clearly evolve easily, but their fate in nature is uncertain. Rare cheaters exploit the wild-type bacteria to increase in frequency within groups. But when they become common, they disrupt sporulation, and the whole group suffers¹. So, an understanding of population struc-

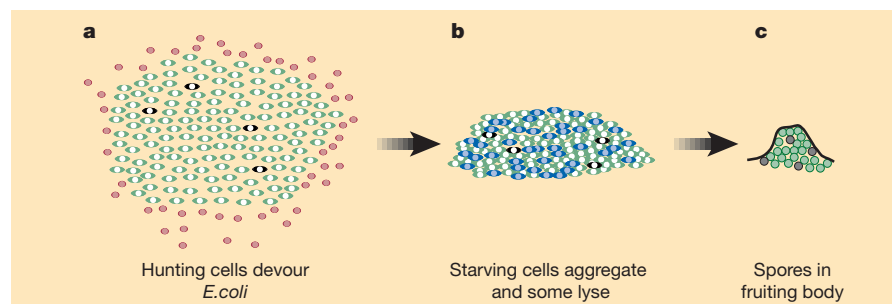


Figure 1 Cheating bacteria. a, Hunting *Myxococcus xanthus* individuals (wild-type are shown in green and cheaters in black) eat *Escherichia coli* (red) by secreting enzymes that degrade their prey. b, In the absence of prey, starving individuals aggregate even more closely into groups of 10^5 cells; many lyse (blue), whereas others form spores. c, Hardy spores in a fruiting body (end-on view) await more favourable conditions. Velicer *et al.*¹ show that cheaters, or developmental mutants, have increased in frequency at this stage because they sporulate more efficiently than the wild-type bacteria.

ture is crucial in comprehending the evolutionary importance of cheaters. If each group of *M. xanthus* is a clonal descendant of a single cell — like the cells in a fruitfly — then cheater groups will lose out to wild-type groups. If, in contrast, *M. xanthus* groups are mixtures of genetically distinct clones — like a pride of lions — then individual cheater mutants should be common in populations. In this case, the myxobacteria fall short as a developmental model for higher organisms, because between-cell signalling will have competitive elements not present in the development of organisms that have passed through a single-cell bottleneck⁶. Groups of myxobacteria would therefore be more analogous to social groups, with all of their inherent potential for conflict.

Velicer *et al.* argue that genotypic diversity within *M. xanthus* fruiting bodies is likely

to be common in the wild. But this will need to be confirmed before we can decide whether groups of these bacteria are more like the body of a fruitfly or a pride of lions — a model for development or for social evolution. ■

Joan E. Strassmann is in the Department of Ecology and Evolutionary Biology, Rice University, PO Box 1892, Houston, Texas 77251, USA.
e-mail: strassm@rice.edu

1. Velicer G. J., Kroos, L. & Lenski, R. E. *Nature* **403**, 598–601 (2000).
2. Dworkin, M. & Kaiser, D. *Myxobacteria* Vol. II (Am. Soc. Microbiol., Washington DC, 1993).
3. Dworkin, M. *Microbiol. Rev.* **60**, 70–102 (1996).
4. Rosenbluh, A. & Rosenberg, E. in *Myxobacteria* Vol. II (eds Dworkin, M. & Kaiser, D.) 213–233 (Am. Soc. Microbiol., Washington DC, 1993).
5. Velicer, G. J., Kroos, L. & Lenski, R. E. *Proc. Natl Acad. Sci. USA* **95**, 12376–12380 (1998).
6. Grosberg, R. K. & Strathmann, R. R. *Trends Ecol. Evol.* **13**, 112–116 (1998).

Magnetoresistance

A new spin on magnets

Thomas F. Rosenbaum

Just like photons propagating through space, electrons flowing through a crystalline metal travel in simple plane waves. Add sufficient disorder and those electrons are no longer free to fly. Rather they will diffuse from scatterer to scatterer, performing a random walk through the 'dirty' metal. At very low temperatures, near absolute zero, the phase of the electron becomes important and quantum interference effects manifest themselves. At the quantum level both the amplitude and phase of the electronic wavefunction are needed to describe the electron completely. On page 581 of this issue, Manyala *et al.*¹ describe how these quantum interference effects not only dominate the behaviour of a metallic ferromagnet (a material similar to iron) at the relatively scorching temperature of liquid nitrogen, but might account for an entirely

new mechanism behind the magnetic response of solids.

In most models of charge transport in metals, it is safe to ignore the interactions between electrons. The Pauli exclusion principle (which forbids two electrons from being in the same state), and the interaction between negatively charged electrons and the positively charged ions in the lattice, dominate the thermal, optical and electrical properties. Electronic diffusion through a disordered medium such as a metallic alloy or a doped semiconductor — materials essential to the electronics industry — permits the charge carriers to interact more strongly (Fig. 1). The effects of electron–electron interactions in the presence of disorder are most pronounced at very low temperatures, generally within a degree of absolute zero, where the relative phases of the

diffusing electrons are not scrambled by thermal fluctuations².

Electrons carry spin as well as charge. Apply a magnetic field and spin-up electrons acquire a different energy from spin-down electrons. The energy difference created by the spins, combined with quantum interference effects, leads to an electrical resistance that changes with the square root of the applied magnetic field. This effect, known as magnetoresistance, has attracted much interest because of its technological potential, for example in magnetic data storage. Many different materials have shown magnetoresistance that involves scattering of conduction electrons by magnetic electrons. For example, when a magnetic field is applied the scattering may be reduced, leading to negative magnetoresistance.

In the ferromagnet studied by Manyala *et al.*¹, the unusual square-root dependence of the electrical resistance on magnetic field, and the positive nature of the magnetoresistance, point to a different underlying mechanism. In fact, the magnetoresistance of a classic semiconductor such as phosphorus-doped silicon, in which the effect was first observed³, switches from negative to positive when quantum interference effects emerge as the temperature drops below 1 K. This means that the resistance switches from shrinking with magnetic field to growing with magnetic field as the temperature is lowered.

Remarkably, Manyala *et al.*¹ show that the disordered, metallic ferromagnet $\text{Fe}_{1-y}\text{Co}_y\text{Si}$ behaves just like a doped semiconductor, but at a temperature two orders of magnitude higher. It is the large (megagauss) internal fields of the ferromagnet that allow this behaviour to occur at higher temperatures. The same electrons, with intertwined electronic spin and charge, are responsible for both the magnetic and electrical properties of $\text{Fe}_{1-y}\text{Co}_y\text{Si}$. This is in contrast to what happens in most magnetoresistive materials, where the electrons involved in electrical conduction are different from those responsible for the magnetism. In $\text{Fe}_{1-y}\text{Co}_y\text{Si}$, the effects of quantum coherence have been amplified by the magnetic nature of the doped FeSi host, in the same way that shrinking electronic devices to the nanometre scale reveals the subtleties of quantum processes at room temperature in quantum dots⁴ and magnetic heterostructures⁵.

The consequences of magnetoresistance in $\text{Fe}_{1-y}\text{Co}_y\text{Si}$ extend beyond the new insights that it provides into the fundamental quantum nature of charge transport in magnets. An electrical resistance that changes with magnetic field is the primary means of information storage in the hard disk on a computer. Magnetoresistive sensors are also found at the business end of a car's speedometer. New materials continue to emerge as potential candidates for mag-

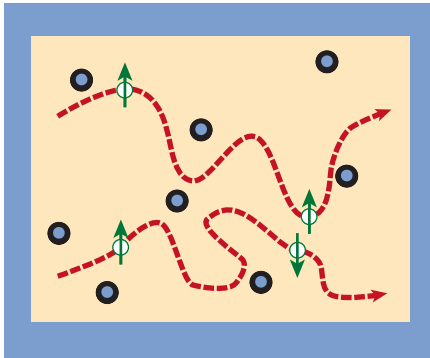


Figure 1 Electrons diffusing through a disordered metal have more opportunity to interact. These interactions can be different for spin-up and spin-down electrons in a ferromagnet. According to Manyala *et al.*¹ these interactions provide a new mechanism for magnetoresistance, the effect used to store information on a computer's hard disk.

netoresistive devices. These include artificially engineered nanostructures in which interfacial spins modulate the electron transport, producing 'giant' magnetoresistance⁶, as well as manganite perovskites with metal-insulator transitions driven by magnetic fields, commonly referred to as 'colossal' magnetoresistance⁷.

The magnetoresistance of $\text{Fe}_{1-y}\text{Co}_y\text{Si}$ is relatively modest, but it results from a new

microscopic mechanism that could lead to the development of different magnetic materials. Moreover, it suggests that there is a universal behaviour underlying the properties of an entire class of complex materials. Here, the parent compound, FeSi, is basically an insulator with peculiar electrical, magnetic and optical properties. But when it is doped, it behaves just like a simple doped semiconductor with enhanced interactions between electrons⁸. Thinking about FeSi in these terms helps to explain why a compound that is 50% iron is not a magnet, whereas adding a non-magnetic dopant such as cobalt creates a metallic ferromagnet. A host of tailored materials, and no doubt some scientific surprises, should be possible through a combination of physical insight and the richness of the periodic table.

Thomas F. Rosenbaum is in the Department of Physics and the James Franck Institute, University of Chicago, 5640 Ellis Avenue, Chicago, Illinois 60637, USA.

e-mail: tfr@rainbow.uchicago.edu

1. Manyala, N. *et al.* *Nature* **404**, 581–584 (2000).
2. Lee, P. A. & Ramakrishnan, T. V. *Rev. Mod. Phys.* **57**, 287–337 (1985).
3. Rosenbaum, T. F. *et al.* *Phys. Rev. Lett.* **47**, 1758–1761 (1981).
4. Ashoori, R. C. *Nature* **379**, 413–419 (1996).
5. Kikkawa, J. M., Smorchkova, I. P., Samarth, N. & Awschalom, D. D. *Science* **277**, 1284–1287 (1997).
6. Jin, S. *et al.* *Science* **264**, 413–415 (1994).
7. Tokura, Y. *et al.* *J. Appl. Phys.* **79**, 5288–5291 (1996).
8. DiTusa, J. F. *et al.* *Phys. Rev. Lett.* **78**, 2831–2834 (1997).

Neurobiology

Dendrites go up, axons go down

Stephen M. Strittmatter

Dendrites and axons, collectively known as neurites, project from the cell bodies of neurons and are specialized for the receipt and transmission of nerve impulses. During development, both types of neurite must grow with the right spatial orientation to make the correct connections.

On page 567 of this issue¹, a group in the laboratory of Anirvan Ghosh (Polleux *et al.*) describes how a guidance factor called semaphorin 3A (Sema3A) acts as a chemo-attractant for dendrites growing from certain neurons. The same molecule acts as a chemorepellant for axons growing from these neurons. One notable aspect of the work is that it brings a unifying element to research on axonal and dendritic guidance; another is that it highlights the role of cyclic nucleotides in creating nervous system patterning.

Numerous axon guidance factors have been identified and classified as attractive or repulsive (or sometimes both)². Among them are the semaphorins, a large family of secreted and membrane-associated factors, one of which (Sema3A) repulses primary

sensory axons in the peripheral nervous system during early development³.

Ghosh's group is exploring the function of Sema3A in the development of cerebral cortex. They work on pyramidal neurons, which carry information from the cerebral cortex to various target areas elsewhere in the brain and spinal cord. In earlier research⁴, they labelled cortical neurons, plated them on slices of cortex, and found that this slice-overlay technique faithfully recapitulated the growth trajectories of pyramidal cell axons *in vivo*. Other evidence⁴ — from co-culture experiments, antibody perturbation studies and analyses of mice in which the gene encoding Sema3A had been knocked out — implicated Sema3A and its receptor subunit neuropilin-1 (NP-1) in directing axon projections. The data suggested that there is a gradient of Sema3A in the developing cortex; its concentration is high in one region (near the pial surface) and low in another (the subcortical white matter). The implication was that axons are repelled by Sema3A and grow down the gradient towards the subcortical zone.

Polleux *et al.*¹ have now extended the analysis of cortical slice overlays to look at dendrite orientation. Pyramidal cell dendrites grow from the apical side of the cell body, opposite the axons. They develop after axons, and extend in the opposite direction, towards the pial surface. Remarkably, it turns out that Sema3A and NP-1 are also required for dendritic patterning. An altered Sema3A gradient can redirect apical dendrites, and dendrite orientation does not depend on previously established axonal orientation. These results were confirmed by an analysis of mouse cortex lacking Sema3A. Thus, the apical dendrite of cortical pyramidal cells extends up the Sema3A gradient towards the pial surface even though the axon of the same cell is repelled by Sema3A (Fig. 1).

The Sema3A gradient seems not only to direct apical dendrite extension but also to determine the site at which the dendrite forms on the cell body. Similarly, studies of semaphorin action in the grasshopper limb bud show that semaphorins contribute to the place of neurite initiation as well as directing the growth of established neurites⁵.

Although other semaphorins have been reported to attract certain neurites^{6,7}, this is the first example of Sema3A doing so. More significantly, it is the first instance of two neurites from the same cell responding in opposite fashion to the same guidance signal. A concentration gradient formed by one

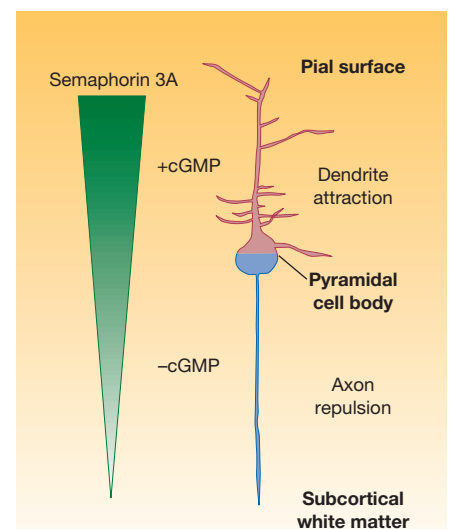


Figure 1 Development of axonal and dendritic architecture in the cerebral cortex. Pyramidal cells in the cerebral cortex project a dendrite from the apical surface of the cell body towards the pial surface; an axon grows in the opposite direction, towards the subcortical white matter. Polleux *et al.*¹ show that a gradient of semaphorin 3A (Sema3A) in the developing cortex is responsible for the orientation of both the dendrite and the axon. An asymmetric concentration of soluble guanylyl cyclase, which creates the signalling molecule cGMP, might underlie dendrite attraction to Sema3A.

factor is sufficient to generate opposing axonal and dendritic projections.

What signal transduction mechanisms might be involved in producing these different responses mediated by NP-1? Receptors for semaphorin include the plexins as well as neuropilins^{8,9}. Plexin distribution in cortical neurons is unknown, but it is conceivable that different plexins associate with NP-1 in axons and in dendrites to produce the opposite guidance response. However, the work of Polleux *et al.* suggests another cause — that the signalling molecule cyclic guanosine monophosphate (cGMP) is a primary determinant of the different responses to *Sema3A*. This possibility was evident from work on isolated *Xenopus* spinal neurons, in which axon responses to *Sema3A* can be converted from repulsion to attraction by raising levels of cGMP¹⁰.

For cortical pyramidal cells, Polleux *et al.*¹ report that soluble guanylyl cyclase is distributed asymmetrically; this is the enzyme responsible for producing cGMP. The asymmetric distribution is first evident as a cap of guanylyl cyclase on the cell body, which precedes the emergence of the apical dendrite itself. How this cap is produced is not known. Nevertheless the high levels of enzyme in apical dendrites imply that local concentrations of cGMP might contribute to attraction. Indeed, reagents that reduce cGMP levels or activity disrupt the dendritic response to *Sema3A*. The axonal response seems to occur in cellular regions of low cGMP concentration and is unaffected by

these reagents. So these studies illuminate one of the first physiological settings in which cyclic nucleotides modulate neurite guidance¹¹.

What next? We need to find out whether other signalling components exhibit differential subcellular localization within various neurons and contribute to selective responses. In addition, other semaphorins (3B and 3C) are expressed in the developing cortex and modify cortical axon outgrowth *in vitro*⁶. It seems that *Sema3A* is a major determinant of both dendritic and axonal trajectories, but the relative contributions of *Sema3B* and 3C to the creation of cortical architecture require investigation. ■

Stephen M. Strittmatter is in the Department of Neurology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520, USA.

e-mail: stephen.strittmatter@yale.edu

1. Polleux, E., Morrow, T. & Ghosh, A. *Nature* **404**, 567–573 (2000).
2. Tessier-Lavigne, M. & Goodman, C. S. *Science* **274**, 1123–1133 (1996).
3. Taniguchi, M. *et al. Neuron* **19**, 519–530 (1997).
4. Polleux, E., Giger, R. J., Ginty, D. D., Kolodkin, A. L. & Ghosh, A. *Science* **282**, 1904–1916 (1998).
5. Isbister, C. M., Tsai, A., Wong, S. T., Kolodkin, A. L. & O'Connor, T. P. *Development* **126**, 2007–2019 (1999).
6. Bagnard, D., Lohrum, M., Uzie, I. D., Puschel, A. W. & Bolz, J. *Development* **125**, 5043–5053 (1998).
7. Wong, J. T., Wong, S. T. & O'Connor, T. P. *Nature Neurosci.* **2**, 798–803 (1999).
8. Takahashi, T. *et al. Cell* **99**, 59–69 (1999).
9. Tamagnone, L. *et al. Cell* **99**, 71–80 (1999).
10. Song, H. *et al. Science* **281**, 1515–1518 (1998).
11. Hopker, V. H., Shewan, D., Tessier-Lavigne, M., Poo, M. & Holt, C. *Nature* **401**, 69–73 (1999).

Ornithology

British birds by number

Robert M. May

So far, about 1.5 million different species of plants and animals — or, more strictly, eukaryotes — have been named and recorded. No one really knows how many different eukaryote species may be alive on Earth today; defensible estimates range from 3 million to 100 million or more.

But we know a lot about birds. They tend to be around during the day, fairly conspicuous, and often pretty or doing interesting things (or both). Globally, the recorded total of living species, close to 10,000, is almost surely within a couple of per cent of the true

total. And we know a lot about British birds in particular. Britain is a small, densely populated country, with a long tradition of public enthusiasm for natural history (especially from energetic clergymen). One contemporary measure of British affection for things avian is that the Royal Society for the Protection of Birds (RSPB) has around one million members, whereas hedgehogs (top of the non-domesticated mammals) only attract around 10,000 friends. There are no societies to express public goodwill towards nematodes.

So it is not surprising that when in May 1999 Britain's Department of Environment, Transport and the Regions (DETR) issued its policy document, *A Better Quality Of Life*, among its 14 'headline indicators' was one for wildlife based on population trends of common breeding birds. In the latest development, the RSPB and the British Trust for Ornithology (BTO) have just published *The State of the UK's Birds 1999**, the first in an annual series of reports surveying the fates of Britain's bird populations.

The wild-bird headline indicator summarizes the trends in abundance of 139 species over the past 30 years. The observed breeding population of each species is arbitrarily scaled to 100 in 1970, the starting year, and subsequent changes are measured against this; the overall 'common bird index' is a simple average of the 139 species. This index rose to just over 110 in the mid-1970s, as many populations recovered from the severe winters of the early 1960s. It has subsequently fallen about 7% from the mid-1970s to 1998, but is still just over 100.

This figure may look encouraging, but less happy trends lurk within it. Woodland birds (41 of the 139 species) are down 20% from the mid-1970s. Farmland birds (20 species) are down 40%, with intensification of agriculture being widely accepted as the explanation.

In a post-Rio Biodiversity Action Plan, British government agencies identified 25 species of breeding birds that are either globally threatened with extinction or whose populations have halved or worse over the past 25 years, or both. For each such species DETR has published an action plan, with specific targets that require a halt in downward trends, and in most cases a return above mid-1990 levels by 2008. These 25 are subdivided into ten species that are widespread in Britain, and 15 that are scarce or rare.

Of the widespread ten, six (turtle dove, grey partridge, spotted flycatcher, corn bunting, reed bunting, linnet) showed continuing decreases during the late 1990s, while four (bullfinch, skylark, song thrush, tree sparrow) can be read as roughly holding their own. On the trends illustrated in the RSPB/BTO graphs, none shows signs of achieving its target for 2008.

The 15 scarce or rare species are a more mixed bunch. With some, there are grounds for mild optimism (stone curlew, red-necked phalarope, corncrake, curlew); with others, there is cause for real pessimism



Figure 1 Under surveillance. Left to right: stone curlew, corncrake, roseate tern, red-backed shrike, marsh harrier and osprey. All photographs by Chris Gomersall, except the red-backed shrike (M. W. Richards).

(roseate tern, red-backed shrike, wryneck). Overall, populations of scarce or rare species have, on average, doubled since 1970, although in many cases this is a reflection of the low numbers in 1970.

Some cheerful news is to be found among British birds of prey. Five species of raptor, put on the *Red List* of endangered species because of historical declines caused by human persecution, are rising in numbers. These are real success stories, which stem from protection of nests against illegal egg collectors and, in some cases, reintroductions (all of the following are numbers of pairs). Red kites have recovered from around 30 in 1970 to 295 in 1999; white-tailed eagles from 0 to 18; marsh harriers from virtually 0 to 147 in 1997; ospreys from less than ten to 130 in 1998; and merlins from 600 in 1984 to 1,300 in 1994. Conversely, the number of pairs of hen harriers has fallen from 534 in 1989 to 521 in 1998; this, the RSPB/BTO report emphasizes, is "due to the illegal persecution associated with management for driven grouse shooting".

Looking beyond the 139 species in the common birds index to the total number of breeding bird species in Britain, there has been a modest increase over the past 30 years, from around 200 to around 210. Looking over a larger sweep of time, we find nearly 40 more bird species in Britain today than there were 200 years ago. About one-third of these are deliberate introductions, purple sandpipers, for instance, but the others are natural colonizations (little owls). The causes of this increase are unclear, but are probably a mixture of protection from egg-collectors and climate change — and simply more bird watchers seeking out their objects of desire.

Robert M. May is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

e-mail: robert.may@zoo.ox.ac.uk

* *The State of the UK's Birds 1999* by R. D. Gregory, D. G. Noble, L. H. Campbell and D. W. Gibbons. Available at no charge from Dr R. D. Gregory, RSPB, The Lodge, Sandy, Bedfordshire SG19 2DL, UK. <http://www.rspb.org.uk>

Cell cycle

A new check on issuing the licence

J. Julian Blow and Shusuke Tada

It is crucial for the continuation of life that, as a cell replicates and divides, the genetic information encoded in its DNA (its genome) is duplicated and passed on accurately to the two new daughter cells. All of the genome must be replicated, with nothing left unreplicated and nothing replicated more than once. This apparently simple problem is made complicated by the way in which eukaryotic cells (those with a defined nucleus) go about copying their large genomes. 'Replication forks' progress along double-stranded DNA, copying both strands at a rate of 10 to 100 base pairs per second. At this rate, a single replication fork would take years to duplicate a typical genome. So, to speed things up, eukaryotes initiate replication forks at thousands of different places — called replication 'origins' — scattered throughout the genome.

This raises a new problem, however: replication forks must never be initiated on DNA that has already replicated during that cell-division cycle. Although general principles have been proposed to explain how cells prevent such re-replication, the precise mechanism is unknown. Writing in this issue, Nishitani and colleagues¹ and Maiorano and co-workers² now suggest the importance of a protein known as Cdt1 in this process.

Previous experiments have shown that precise DNA duplication is achieved by sepa-

rating the initiation of replication into two phases³. In the first phase, which occurs early in the cell cycle, replication origins are 'licensed' by becoming loaded with the RLF-M (for replication-licensing factor type 'M') complex of MCM proteins⁴ (also known as MCM/P1 proteins). In the second phase, an activity that promotes the DNA-replication (S) phase of the cell cycle — S-phase-promoting factor, or SPF — swings into action. SPF initiates a single pair of replication forks at each licensed origin. As replication forks proceed from the origin, MCM proteins are removed, thereby resetting the origin to the 'unlicensed' state. To ensure that initiation occurs no more than once at each origin, licensing activity must cease completely before SPF becomes active. Even a small amount of licensing activity present during S phase could lead to the re-replication of part of the genome. According to the new papers^{1,2}, the Cdt1 protein acts in the tightly controlled pathway that promotes licensing only early in the cell cycle.

The *cdt1* gene was initially identified in the fission yeast *Schizosaccharomyces pombe* as a gene that is essential for DNA replication. It is periodically expressed under the control of the transcription factor Cdc10 (ref. 5). These properties are shared by the *cdc18* gene in *S. pombe* and its homologue, *CDC6*, in other eukaryotes. Cdc6/Cdc18 is essential for licensing replication origins⁶

(Fig. 1). Overexpression of Cdc6/Cdc18 in *S. pombe* causes DNA to replicate several times within a single cell cycle^{7,8}, probably as a consequence of keeping the licensing system active late in the cycle, after DNA replication has started.

Nishitani *et al.*¹ now show that Cdt1 has a similar function to Cdc6/Cdc18, and is required for origin licensing. They show that, in cells in which Cdc6/Cdc18 is only slightly overexpressed (at levels that alone do not induce significant re-replication of DNA), coexpression of Cdt1 causes high levels of re-replication. The inability of overexpressed Cdt1 to induce over-replication by itself indicates that it might need to be less tightly repressed later in the cell cycle than is Cdc6/Cdc18. In a parallel series of experiments, Maiorano *et al.*² cloned the *Xenopus laevis* (frog) *CDT1* gene, and investigated its function in cell-free extracts that support

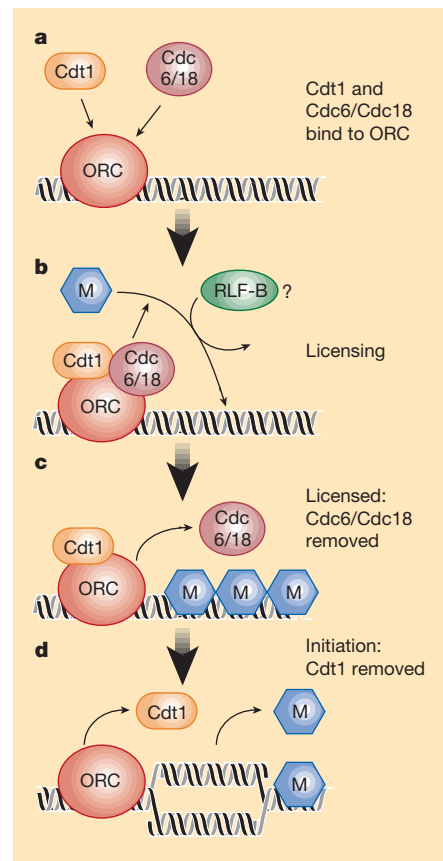


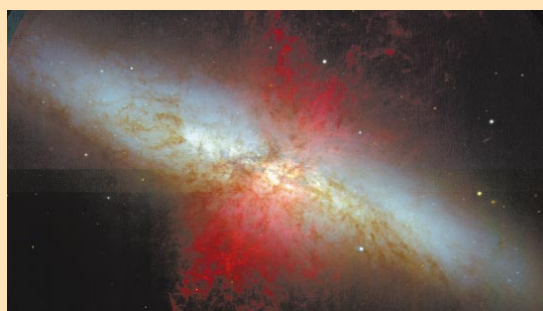
Figure 1 Steps in the licensing of a DNA-replication origin. a, Cdt1 (as shown by Nishitani *et al.*¹ and Maiorano *et al.*²) and Cdc6/Cdc18 bind to the origin-recognition complex (ORC) at an origin of DNA replication. b, Several MCM protein complexes, represented by 'M', are then assembled onto the origin to 'license' it for replication. The activity of replication-licensing factor B (RLF-B) might be involved at this point. c, As a consequence of replication licensing, Cdc6/Cdc18 is displaced from the origin. d, As DNA replication starts, all of the Cdt1 is removed from the origins. MCM proteins are removed as the DNA replicates.

Astronomy

Two for the price of one

This latest image from Subaru, one of the world's largest telescopes, was taken by the Faint Object Camera and Spectrograph (FOCAS) during its first night of operation in February. The FOCAS instrument is capable of taking two sorts of pictures: images and spectra, which can be chosen at the flick of a switch. Spectra of faint galaxies provide physical information, such as temperature and density, to complement the optical images.

The image here is of the irregular galaxy M82, which is roughly 12 million light years from Earth. The bluish light running across the image comes from stars in the



galactic disk. The red filaments flowing out at right angles are produced by ionized hydrogen gas, which is evidence for star formation at the centre of M82.

Subaru — the Japanese name for the constellation Pleiades — was built by Japan's National Astronomical Observatory at the top of

Mauna Kea in Hawaii. The 8.2-metre telescope saw its first light early in 1999, but its systems are still being tested and fine tuned. Five of the seven observational instruments have now been added and will soon be ready for routine observations.

Sarah Tomlin

cell-cycle progression *in vitro*. The results of Maiorano *et al.*, showing that the *Xenopus* Cdt1 protein is required to license DNA by loading the MCM proteins onto it, are consistent with those obtained in *S. pombe*.

These papers also show that, in both fission yeast and frog, Cdt1 itself associates with DNA. This association is dependent on the presence of the origin-recognition complex (ORC)², as is the association of Cdc6/Cdc18 with DNA⁶. In fission yeast extracts, Cdt1 and Cdc6/Cdc18 can be precipitated together by an antibody against just one of the proteins, suggesting that they might form a complex. However, they can also associate with DNA independently of one another. These results indicate that, early in the G1 phase of the cell cycle (the stage preceding S phase), Cdt1 and Cdc6/Cdc18 may bind together to ORCs on DNA (Fig. 1a), where they are required for licensing to occur (Fig. 1b).

A further essential licensing activity, termed RLF-B, is probably also required at this stage⁹. The protein underlying RLF-B activity is unknown, and could theoretically correspond to Cdt1. However, the behaviour of RLF-B differs from that of Cdt1: unlike Cdt1, RLF-B activity does not associate with unlicensed DNA⁹. Cdt1 also seems to behave differently from Cdc6/Cdc18, as, in the frog system, Cdt1 remains associated with DNA right up until the start of S phase (Fig. 1d), whereas Cdc6/Cdc18 is displaced from DNA immediately after origins are licensed¹⁰ (Fig. 1c). Further studies will be needed before we fully understand the roles of these proteins in origin licensing.

The existence of this complex series of events in origin licensing probably reflects the essential function of this system in maintaining genomic integrity. And it is probably even more important to prevent the re-licensing of replicated origins (and hence the re-initiation of DNA replication) in multicellular organisms than it is in single-celled ones. The occasional re-replication event would have little effect on the proliferative capacity of a strain of single-celled organisms such as yeast. But, by generating a cancer-causing mutation, it might be fatal to a multicelled organism. Re-assembling the Cdt1 protein — which is regulated in different ways (transcriptionally or post-translationally) in yeast¹ and frogs² — at replication origins represents yet another step that is necessary for re-licensing to occur. We can expect the regulation of the replication licensing system to become more complex still.

J. Julian Blow and Shusuke Tada are at the CRC Chromosome Replication Research Group, Department of Biochemistry, University of Dundee, Dow Street, Dundee DD1 5EH, UK. e-mail: j.j.blow@dundee.ac.uk

1. Nishitani, H., Lygerou, Z., Nishimoto, T. & Nurse, P. *Nature* **404**, 625–628 (2000).
2. Maiorano, D., Moreau, J. & Méchali, M. *Nature* **404**, 622–625 (2000).
3. Blow, J. J. & Laskey, R. A. *Nature* **332**, 546–548 (1988).
4. Tye, B. K. *Annu. Rev. Biochem.* **68**, 649–686 (1999).
5. Hofmann, J. F. & Beach, D. *EMBO J.* **13**, 425–434 (1994).
6. Coleman, T. R., Carpenter, P. B. & Dunphy, W. G. *Cell* **87**, 53–63 (1996).
7. Kelly, T. J. *et al.* *Cell* **74**, 371–382 (1993).
8. Nishitani, H. & Nurse, P. *Cell* **83**, 397–405 (1995).
9. Tada, S., Chong, J. P. J., Mahbubani, H. M. & Blow, J. J. *Curr. Biol.* **9**, 211–214 (1999).
10. Rowles, A., Tada, S. & Blow, J. J. *J. Cell Sci.* **112**, 2011–2018 (1999).

Daedalus

The sharper image

Last week Daedalus proposed to use silicon microfabrication to make klystrons with a resonant cavity only a wavelength of light or so across. Such resonant sources could generate visible light directly. For useful brightness, many such cavities would have to be formed on a single silicon chip: perhaps ten thousand per square centimetre. Such a chip could be a wonderful source of intense, coherent, steerable light.

If all the photoklystrons were identical, and spaced as an even array, their outputs would combine in phase to give a coherent, parallel beam. Even the most exact microfabrication could not do this by itself. But a klystron can be fine-tuned by adjusting its electrode voltages, a simple task for the control circuitry integrated on the same chip. With all the klystrons thus phased together, they would form a collective source as parallel and monochromatic as any laser, but far more efficient, and with a useful tunable range.

Even better, by shifting the phase stepwise across the array, the light could be steered at electronic speed, like a radar beam. DREADCO's 'Photophasor' lamp could give 'lidar' the fast scanning power of the best radar. But its most important use will surely be in large-scale imaging. Driven by standard TV scanning circuitry, the Photophasor will be able to draw a bright moving image on a big screen. With red, green and blue klystrons spaced on the chip like the tricolour matrix of dots on a television screen, it could write a big bright full-colour motion picture under direct digital control.

Electronic cinema and big-screen display will be transformed. The clumsy, blurry cathode-ray-tube projector will be swept away by the simple, elegant Photophasor. Highly efficient, it will be a cool, low-power device, interfacing directly with TV or computer inputs. Cinemas will abandon film for DVD disks screened by Photophasor, and Photophasor business presentations will have even more compulsive visual gorgeousness to disguise their lack of ideas. And by scaling down the technology, a simple Photophasor display could even displace the ubiquitous cathode ray TV and computer screen. The crippling limit of the personal computer, the low resolution and limited area of its display, would be overcome at last. **David Jones**

The Further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special *Nature* offer: m.curtis@nature.com

Obituary

George Ledyard Stebbins (1906–2000)

George Ledyard Stebbins, whose synthesis of genetics and evolutionary theory from the plant world was central to formulating the modern synthesis of evolution, died on 19 January at the age of 94. As well as being a renowned researcher, he was a prolific writer, a popular teacher and an outspoken conservationist.

One hundred years ago, Hugo de Vries, Carl Correns and Eric von Tschermak independently rediscovered Gregory Mendel's insight that each parent contributes to each segregating character only one genetic unit, later called the 'gene'. Historians of science designated this epoch-making event the 'rediscovery of Mendel' because it led to the birth of genetics, a powerful new science that transformed the biological sciences in the twentieth century. Within only two decades of the rediscovery, most of the major principles of the transmission of genetic traits were worked out by an army of converts to the science.

Unfortunately, advocates of the new science such as de Vries believed that genetic changes or 'mutations' were discontinuous and large-scale. These theories of evolution by sudden, large changes undermined darwinian theories, which instead stressed gradual evolution through natural selection operating on small, individual differences. As a result, advocates of darwinian evolution, including many natural systematists who studied geographical variation and were aware of populational variation, turned away from mendelian genetics. What ensued was what Ernst Mayr subsequently described as a "deplorable communication gap" between two opposing "camps of biologists".

This gap began to close with the integration of mendelian genetics and darwinian evolution during the period 1920–1950. First through the work of mathematical population theorists such as R. A. Fisher, J. B. S. Haldane and Sewall Wright, and then through the coordinated efforts of Theodosius Dobzhansky, Bernhard Rensch, Ernst Mayr, Julian Huxley and George Gaylord Simpson, a view of evolution emerged in which major evolutionary phenomena including speciation, the origin of evolutionary variation, large-scale evolutionary trends and systematic hierarchy could be understood in terms of genetics. Natural selection was restored as the primary mechanism responsible for evolution.

Most of the leading 'architects' of the



Botanical architect of the evolutionary synthesis

'evolutionary synthesis' were zoologists. Only one was a botanist who concentrated on understanding variation and evolution in plants. This was George Ledyard Stebbins. In 1950 Stebbins brought botany into the evolutionary synthesis through his own monumental book, *Variation and Evolution in Plants*. It served a dual purpose: not only did it demonstrate how evolutionary mechanisms operated in plants at the genetic level, thereby bringing plant evolution into line with animal evolution as it emerged from Dobzhansky's 1937 *Genetics and the Origin of Species*, but it also organized a disparate set of disciplines into a new field, plant evolutionary biology. Stebbins's book was the longest and the last of the works that formed the evolutionary synthesis and is one of the most influential books in the history of modern botany. From 1950 onwards, Stebbins was the botanical architect of the evolutionary synthesis.

Stebbins was descended from a wealthy New England family, which supported his interest in natural history. He began his botanical career at Harvard University, studying taxonomy, biogeography and cytology. After an appointment at Colgate University in 1931, Stebbins joined the University of California, Berkeley, in 1935 as a junior geneticist. In 1950 he left Berkeley for the Davis campus, where he was instrumental in founding the genetics department. He remained there until his death and was one of the most celebrated figures at the university.

Stebbins's knack for synthesis accompanied life-long wide-ranging interests and an unusual career path. Whereas other botanists of his generation were specialists who concentrated on one plant group, Stebbins worked on a staggering number of organisms, including grasses, peonies and the latex-producing plant guayule. His contributions began in floristics (the taxonomic study of flora) and then progressed much as biology did in the twentieth century to include systematics, cytogenetics, plant evolutionary biology, plant breeding and, finally, developmental and molecular biology. He was a laboratory worker, microscopist, agriculturist, theorist, keen naturalist and even a pioneering conservationist. In addition to original research articles and scientific monographs, he wrote many influential textbooks of general biology and evolution, and semi-popular books. He excelled at the technical review article and included an extraordinary number in his long bibliography.

Stebbins had much in common with Mayr, Huxley and the other architects: all had similar wide-ranging interests and diverse careers. All were voracious readers with an ability to see the larger picture and all were prolific writers. Through the 1960s and the 1970s, all became public champions of evolution. Stebbins actively fought scientific creationism in the United States, and echoed Dobzhansky's famous assertion that "nothing in biology makes sense except in the light of evolution".

All who had contact with Stebbins became aware of his endless energy and infectious enthusiasm. He had an exceptional memory with an agile and curious intellect. He was unafraid of provocative ideas or unorthodox interpretations. He was exceptionally articulate and spoke with a crisp New England accent in perfect paragraphs; often he both delighted and annoyed colleagues with his tendency to dominate conversations or to break out unexpectedly into funny verse or song. His temper tantrums became the topic of amused conversation and one incident involving a typewriter thrown out of a window became the stuff of urban legends.

Even at an advanced age, Stebbins took an active interest in following scientific developments. Right up to his death, he demonstrated a resilience and love of the subject more characteristic of younger people.

Vassiliki Betty Smocovitis

Vassiliki Betty Smocovitis is in the Department of History, 4131 Turlington Hall, University of Florida, Gainesville, Florida 32611, USA.
e-mail: bsmocovi@history.ufl.edu

Payment for labour in monkeys

Capuchins will voluntarily share treats with other monkeys that helped to secure them.

Cooperative hunting, in which several individuals pursue prey but only one makes a capture, is central to theories of human social and moral evolution^{1–3}. But among other primates, it is known only from the chimpanzee and a large-brained neotropical monkey, the capuchin^{4–7}. It probably evolved through either mutualism, in which two or more cooperators benefit simultaneously, or reciprocal altruism, in which one favour is repaid by another^{8,9}. We have found that brown capuchins (*Cebus apella*) share rewards obtained by a joint effort more readily than rewards obtained individually. Even if hunting in the field involves selfish opportunism, this

food incentive will greatly enhance the persistence of cooperation.

We made a pair of monkeys work for food¹⁰, by placing two transparent bowls in full view. Each bowl was accessible to one monkey by pulling the tray towards itself using one of two protruding bars (Fig. 1a). Individual strength was periodically tested to determine each individual's maximum pulling weight. We investigated three conditions: solo effort (only one monkey had access to a pull bar and a baited cup, and the tray was counterweighted within this individual's pulling capacity); cooperation (both monkeys had pull bars, and the tray was counterweighted such that the strength of both was required, but only one bowl was baited), and mutualism (same as cooperation, but with both bowls baited).

We put eight apple slices into the bowl(s) at the start of each of four 10-min trials per test. After pre-training, we did a minimum of 24 cooperation, 8 mutualism and 8 solo-effort tests on each same-sex pair (5 female and 2 male pairs) of unrelated adults from the same social group, applying half the tests to each direction within a pair. The capuchins successfully pulled in the tray in a mean ($\pm$ s. e.) of $85.4 \pm 3.5\%$ of solo-effort trials and $88.9 \pm 2.7\%$ of mutualism trials. The success rate for cooperation trials was substantially lower, $39.2 \pm 3.1\%$ (paired comparison with mutualism: $t = 14.06$, $P < 0.001$; with solo effort: $t = 14.56$, $P < 0.001$).

Capuchins will share attractive foods spontaneously, even if separated by a mesh restraint — a pattern known as “facilitated taking”, with the possessor approaching the divider and dropping crumbs or whole pieces while the partner reaches for the food¹¹. As the possessor could monopolize the food by avoiding the divider, both parties play an active role. Facilitated taking is reciprocal across individuals as well as across time between any two individuals¹².

We measured the amount of sharing as the number of times the partner's hand reached through the mesh to pick up food from the other side, limiting ourselves to unambiguous videotaped behaviour. We ignored mutualism tests, in which each individual had its own food. Compared with solo-effort tests, significantly more pieces of food were shared after successful cooperation trials. Moreover, a greater proportion of transfers after cooperation were of a tolerant nature (Fig. 1b).

Because the individual with the rewarding bowl was invariably motivated to pull, the helper's behaviour was decisive. We



Figure 2 A hopeful young capuchin watches an adult eating.

found that the helper pulled two to three times more often in cooperation trials if the preceding trial had been successful than if it had been a failure (analysis of variance: $F_{1,11} = 12.41$, $P = 0.003$, directional). This suggests either a stable motivation across trials or a causal connection between a share of the reward (9 out of 10 successful trials resulted in food transfer) and subsequent willingness to pull.

We have shown that capuchins cooperate even if it is obvious that only one of them, and which one, will be rewarded. The increase in sharing following cooperation may rest on psychological mechanisms as complex as mental score-keeping of services¹³ and “gratitude”, or as simple as attitudinal reciprocity¹². According to the latter explanation, a joint effort, and the mutual coordination this entails, may induce a positive attitude towards the partner, reflected in attraction and social tolerance. If this facilitates the sharing of pay-offs, in turn providing an incentive for continued cooperation, we have two mechanisms that together function as payment for labour and labour for payment.

Frans B. M. de Waal, Michelle L. Berger
Living Links, Yerkes Regional Primate Research Center, and Department of Psychology, Emory University, Atlanta, Georgia 30329, USA
e-mail: dewaal@emory.edu

1. de Waal, F. B. M. *Good Natured* (Harvard Univ. Press, Cambridge, MA, 1996).
2. Whiten, A. in *The Descent of Mind* (eds Corballis, M. C. & Lee, S. E. G.) 173–193 (Oxford Univ. Press, 1999).
3. Stanford, C. B. *The Hunting Apes* (Princeton Univ. Press, Princeton, NJ, 1999).
4. Goodall, J. in *Primate Behavior* (ed. DeVore, I.) 425–473 (Holt, New York, 1965).
5. Boesch, C. & Boesch, H. *Am. J. Phys. Anthropol.* **78**, 547–573 (1989).
6. Perry, S. & Rose, L. *Primates* **35**, 409–415 (1994).
7. Rose, L. *Int. J. Primatol.* **18**, 727–765 (1997).
8. Trivers, R. L. *Q. Rev. Biol.* **46**, 35–57 (1971).
9. Dugatkin, L. A. *Cooperation Among Animals* (Oxford Univ. Press, 1997).
10. Crawford, M. P. *Comp. Psych. Monogr.* **14**, 1–88 (1937).
11. de Waal, F. B. M. *J. Comp. Psychol.* **111**, 370–378 (1997).
12. de Waal, F. B. M. *Anim. Behav.* (in the press).
13. de Waal, F. B. M. *Evol. Hum. Behav.* **18**, 375–386 (1997).

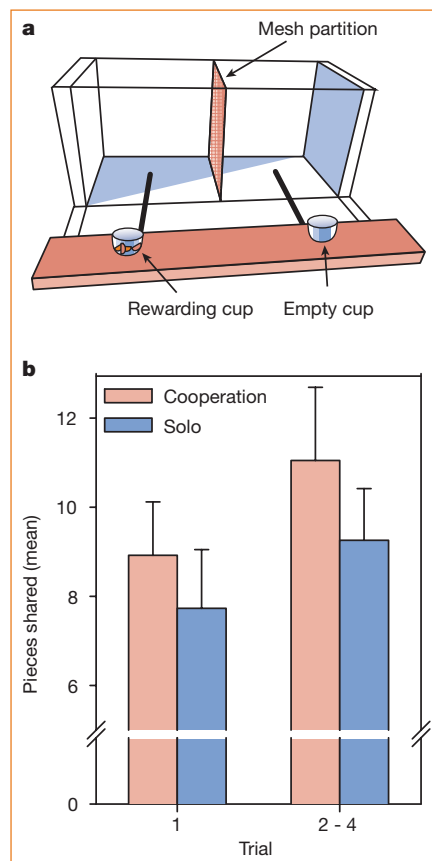


Figure 1 Cooperation and sharing. **a**, In a $144 \times 60 \times 60$ cm test chamber, two monkeys were divided by a mesh partition. In cooperation tests, the strength of both monkeys was required to pull in a tray with two transparent bowls. In all 4 trials per test, apple slices were placed in the same bowl. In solo-effort controls, the bar in front of the empty cup was removed and the counterweight reduced. **b**, The rate of facilitated taking was higher after cooperation than after solo controls (ANOVA: $F_{1,12} = 5.63$, $P = 0.018$, directional). Also, the mean ($\pm$ s. e.) percentage of food collections taking place within reach and sight of the possessor was $64.8 \pm 4.5\%$ after cooperation, compared with $57.9 \pm 4.8\%$ after solo controls ($F_{1,12} = 10.44$, $P = 0.0035$, directional). For further details see www.emory.edu/LIVING_LINKS/

Reproductive systems

Sex and the single lichen

Lichens are physiologically adapted for growth in dry, nutrient-deficient and temporarily thermally extreme habitats¹, but it is unclear how the reproductive strategies of lichen symbionts might have evolved to maximize colonization. We now show that sexually reproducing lichen-forming fungi can self-fertilize, and propose that this breeding system allows these symbiotic organisms to reproduce successfully in harsh environments.

Most lichenized fungi produce abundant sexual structures, and in many species sexual spores seem to provide the only means of dispersal. For example, 90% of lichens found in Great Britain and Ireland² produce ascogonia (fruit bodies) containing sexually derived ascospores, whereas only 29% form symbiotic vegetative propagules. Sex in lichenized fungi has been assumed to equate with outcrossing³, but failure to induce sexuality *in vitro* has prevented experimental investigation of their breeding systems.

We avoided this problem by using molecular markers to elucidate the sexual cycle. We compared the DNA fingerprints of single-spore progeny from a single ascogonium: the presence or absence of genetic variation

between sibling spores would show whether outcrossing (heterothallism) or self-fertilization (homothallism: allowing the fusion of two identical nuclei during meiosis) had occurred.

The crustose lichens *Graphis scripta* (order: Ostropales) and *Ochrolechia parella* (order: Pertusariales) fruit abundantly, but neither produces symbiotic vegetative propagules (Fig. 1a,b). We collected discrete, symmetrical lichen thalli at locations more than 10 m apart in south Wales, and induced excised individual ascogonia to discharge spores. Extraction of DNA from cultured mycelia derived from single spores gave 218–263 randomly amplified polymorphic DNA (RAPD) markers⁴ for each isolate. In both species, these markers revealed that spores from the same ascogonium are genetically uniform (Fig. 1c), providing compelling evidence of homothallism.

This breeding system probably serves an ecological function similar to that of self-pollination in flowering plants^{5–7}. It may confer a selective advantage by facilitating high spore output without the need for outcrossing; it would also promote the development of a lichen population from a single pioneer spore after dispersal into a new site. Genetic stability may be advantageous in abiotically extreme but relatively undisturbed habitats in which there is a low intensity of biotic interactions because it

perpetuates successful genotypes adapted to the prevailing environmental regimes. Homothallism also retains certain benefits of sexual over asexual reproduction, including opportunistic outcrossing, as obligate selfing is rare in homothallic fungi⁸.

Although we found both lichen fungi to be homothallic, the ascospore offspring derived from different conspecific thalli were genetically distinct (Fig. 1d). The average RAPD genetic divergence⁹ observed in five isolates of *G. scripta* from one woodland was 15.2% (according to a neighbour-joining analysis of Jaccard's coefficient of band matching). Furthermore, ascospore progeny from different ascogonia on 'single' thalli of *O. parella* showed RAPD polymorphisms, indicating that these lichen thalli may have been composed of at least two sexually active fungal genotypes.

Comparable variation within individual thalli was not detected in *G. scripta* (Fig. 1d). This difference is probably related to habitat: *G. scripta* is a pioneer, with thalli that probably arise from single spores, whereas the higher rates of spore deposition in the more densely occupied habitats of *O. parella* may cause frequent mergers between developing thalli. These 'mechanical hybridizations' are probably very common¹⁰.

If homothallism is a general characteristic of lichen-forming fungi, including those in the order Lecanorales (which contains the most lichen-forming fungal species), it may explain the ecological paradox of the persistence of fruiting lichens in severe habitats where a stable, highly adapted genetic line would be most advantageous¹¹. It has been shown that sexual reproduction is predominant in lichen communities growing at the frontiers of terrestrial life in polar regions^{12,13}. Lichens may thus provide a model for the evolution of breeding systems in extreme environments.

G. J. Murtagh, P. S. Dyer, P. D. Crittenden

School of Biological Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK
e-mail: pdc@nottingham.ac.uk

1. Kershaw, K. A. *Physiological Ecology of Lichens* (Cambridge Univ. Press, Cambridge, 1985).
2. Purvis, O. W. et al. (eds) *The Lichen Flora of Great Britain and Ireland* (Natural History Museum Publications, London, 1992).
3. Hestmark, G. *Oecologia* **92**, 305–312 (1992).
4. Murtagh, G. J., Dyer, P. S., McClure, P. C. & Crittenden, P. D. *Lichenologist* **31**, 257–267 (1999).
5. Jarne, P. & Charlesworth, D. *Annu. Rev. Ecol. Syst.* **24**, 441–466 (1993).
6. Jain, S. K. *Annu. Rev. Ecol. Syst.* **7**, 469–496 (1976).
7. Richards, A. J. *Plant Breeding Systems* 2nd edn (Chapman & Hall, London, 1997).
8. Burnett, J. H. *Mycogenetics: An Introduction to the General Genetics of Fungi* (Wiley, London, 1975).
9. Weising, K., Nybom, H., Wolff, K. & Meyer, W. *DNA Fingerprinting in Plants and Fungi* (CRC, Boca Raton, 1995).
10. Jahns, H. M. & Ott, S. *Bibl. Lichenol.* **67**, 49–67 (1997).
11. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, Cambridge, 1978).
12. Fahselt, D., Maycock, P. F. & Wong, P. Y. *Lichenologist* **21**, 343–353 (1989).
13. Sancho, L. G. & Valladares, F. *Polar Biol.* **13**, 227–233 (1993).

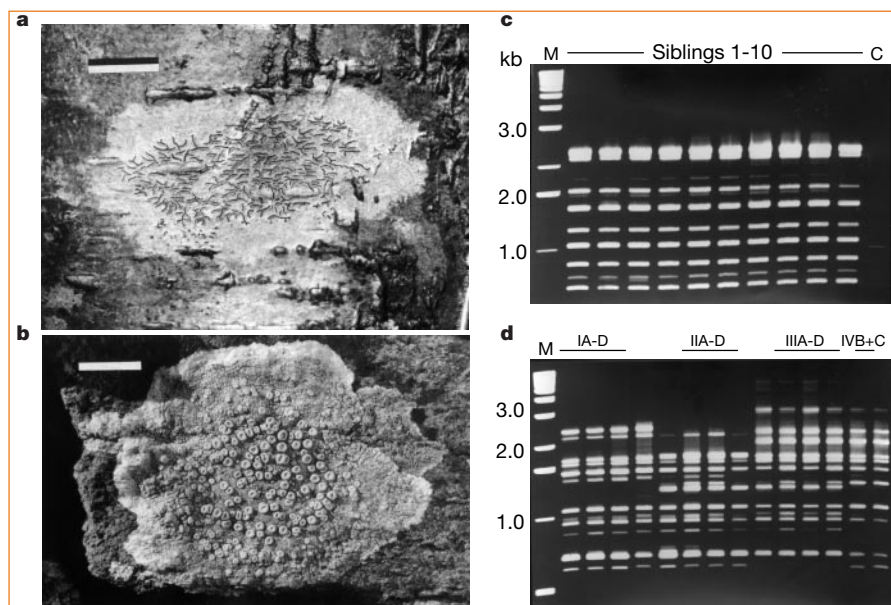


Figure 1 Sexual reproduction and genetic variation in lichen-forming fungi. **a, b**, The crustose lichens *Graphis scripta* and *Ochrolechia parella* growing on tree bark and maritime rocks, respectively (scale bars, 10 mm). Both species develop abundant ascogonia (elongate in *G. scripta* and disc-shaped in *O. parella*), in which sexual ascospores are produced. **c**, Randomly amplified polymorphic DNA (RAPD) markers generated with primer OPAJ-03 showing uniformity in a sample of ten single-spore isolates from one ascogonium of *G. scripta*. A set of ten progeny was collected from each of three separate thalli of both species and assessed for variability in RAPD profiles using a minimum of 30 primers (Operon Technologies). Consistent uniformity among the RAPD markers indicated that for both species all spores from the same ascogonium were genetically identical. **d**, RAPD markers generated with primer OPAX-12 showing polymorphisms between isolates from four different thalli (I–IV) of *G. scripta* but uniformity among spores from 2–4 different ascogonia (replicates A–D) on the same thallus. Eight thalli of *G. scripta* were compared, five of which were collected from the same woodland. For *O. parella*, 3–4 ascogonia on each of three thalli were compared and polymorphisms found between both thalli and ascogonia on the same thallus. M, size markers in kb; C, water control.

Fish behaviour

Stomach rinsing
in rays

The vomiting reflex is a protective mechanism for the bulk ejection of noxious material, and is common in vertebrates¹. Here we show that the thorn-back ray *Raja clavata* (Rajidae) can 'rinse' its stomach by full gastric eversion, washing small indigestible food particles and sloughed gastric mucosa and mucus out of the upper digestive tract. This may be a widespread mechanism, besides vomiting, for fish to remove noxious material from the stomach.

Skates and rays are ubiquitous and comprise nearly half the 900 known species of Elasmobranchii (sharks, skates and rays). We studied the vomiting responses of *R. clavata* caught in the English Channel and fed on chopped fish and squid in the laboratory. Individual *R. clavata* were given an emetic¹, veratrine hydrochloride (10 mg kg⁻¹ i.p. in elasmobranch ringer; *n* = 4), or ringer (*n* = 3). After injection we placed them in a glass-sided tank and recorded their behaviour on video.

Stomach eversion and swallowing of the everted stomach occurred within 10 min of drug injection, but not in rays given ringer. Eversion was frequent: in one animal, it occurred nine times in four discrete bursts post-injection (three eversions between 351 and 361 s; one at 404 s; three between 425 and 430 s; two at 463 s). Immediately before eversion, the mouth was wide open, the jaws protruded and the buccal and pharyngeal cavities expanded (Fig. 1). The stomach prolapsed and was fully everted 0.5–2.0 s later, concomitant with a strong contraction of the pharynx. The outline of the hypo- and basi-branchial cartilages under the skin was visible externally, indicating contraction of the coraco-mandibularis, -hyoideus and -branchiales muscles (see frame 7:08:51).

The entrance to the pylorus on the stomach's mucosal surface was visible, indicating that the stomach was fully everted. This was accompanied by lateral shaking of the head, a behaviour also seen in vomiting dogfish¹. Between 0.2 and 2.0 s later, when the pharyngeal cavity expanded, the stomach was retracted into the mouth. We assume that it was swallowed in the next 0.1–0.2 s, as it was not visible inside the oral cavity, and maximal jaw protrusion and mouth closure followed rapidly — behaviour consistent with the normal prey-capture responses of rays before swallowing. The material ejected ranged from small pieces of fish to a particulate mucoid suspension.

These observations show that rays evert and swallow their stomachs in response to an agent that induces vomiting in other fish (such as dogfish)¹ and vertebrates (such as

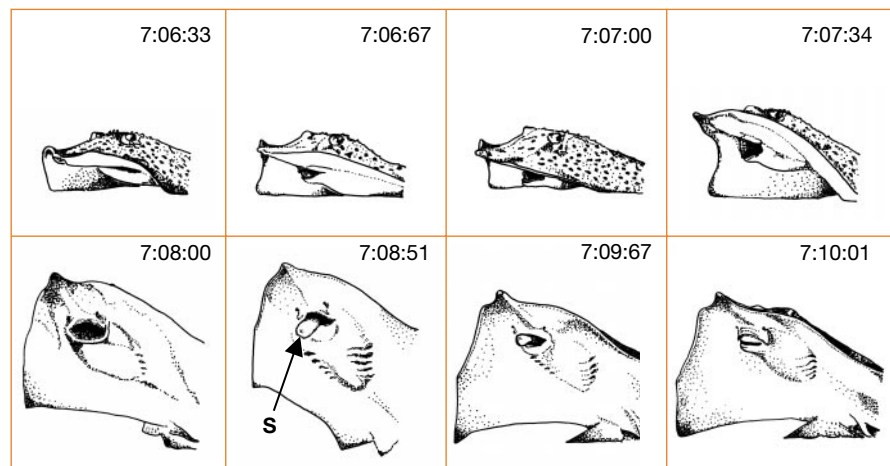


Figure 1 Temporal sequence of stomach eversion and swallowing in *Raja clavata*. Number in each frame refers to elapsed time (minutes: seconds: hundredths of a second) from injection of veratrine. S, cardiac stomach. Video recordings of behaviour were analysed by taking monochrome digitalized still images, constituting individual frames, from the videotape at selected time intervals using a Panasonic NV-MP1 video printer.

crocodiles and cats)^{2,3}. We do not know how the stomach is everted, but it probably involves profound relaxation of the gastric musculature and a rapid increase in intra-abdominal pressure. Some amphibians also prolapse the oesophagus and stomach through the mouth during vomiting^{4,5}, and intense abdominal muscle contraction has been reported as the mechanism for oral eversion in ten species of frogs and toads⁶.

Post-mortem studies of wild-caught rays (*n* = 6) show that oral eversion of the stomach could be induced by gas inflation of the peritoneal cavity using a pressure of about 200 mm Hg. We also found that chitinous food residues, such as prawn exoskeletons, and larger pieces of fish can also be ejected by this procedure.

In rays, the oral eversion of the stomach is not impeded by mesenteric attachments of the gut to the body wall⁷. The short distance from the mouth to the stomach and the short, wide oesophagus — common to both rays and frogs and toads — presumably act to facilitate eversion.

Cloacal protrusion of the valvular intestine in rays could not be induced by peritoneal inflation, although it can be in members of the shark family Carcharhinidae⁸. There are no descriptions of vomiting in these species, but fishermen have observed line- and net-caught sharks with everted stomachs, possibly due to capture; however, post-eversion stomach swallowing has not been recorded⁹.

Nonetheless, there is anecdotal evidence that stomach eversion and swallowing occurs in three carcharhinid shark species (*Carcharhinus cautus*, *Negaprion brevirostris*, *Prionace glauca*) and a guitarfish (*Rhinobatos typus*) (C. Grist, J. Ugoretz and W. White, personal communication). A lemon shark, *N. brevirostris*, was seen to evert its stomach naturally, where it hung limply out of one side of the mouth before

head-shaking released an oily 'scum' from the surface of the stomach, which was then retracted slowly through the mouth and swallowed. The whole episode lasted 25–30 seconds.

We suggest that sharks and rays can externalize most of their alimentary canal, and that this process serves to remove parasites, indigestible material, toxic food, and sloughed gastric mucosa and mucus. The last of these could also be used to excrete accumulated toxic metabolites — as in chinstrap penguins, which excrete fluoride by vomiting their stomach linings¹⁰. In frogs and toads, stomach eversion is accompanied by 'gastric grooming' with the right hand to wipe away vomitus⁵. We propose that rays' head-shaking during oral eversion fulfils a similar function, and rinses the stomach of persistent small particles and mucus.

David W. Sims*, Paul L. R. Andrews†, J. Z. Young‡§

*Department of Zoology, University of Aberdeen, Tillydrone Avenue, Aberdeen AB24 2TZ, UK
e-mail: d.w.sims@abdn.ac.uk

†Department of Physiology, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK

‡Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK

§Deceased

- Andrews, P. L. R., Sims, D. W. & Young, J. Z. *J. Mar. Biol. Assoc. UK* **78**, 1269–1280 (1998).
- Andrews, P. L. R. *et al. J. Exp. Biol.* (in the press).
- Borison, H. L. & Fairbanks, V. F. *J. Pharmacol. Exp. Therapeut.* **105**, 317–325 (1952).
- Naitoh, T. & Wassersug, R. *J. Nature* **380**, 30–31 (1996).
- Naitoh, T. *et al. Physiol. Zool.* **62**, 819–843 (1989).
- Naitoh, T., Imamura, M. & Wassersug, R. *J. Comp. Biochem. Physiol.* **100C**, 353–359 (1991).
- Wake, M. H. (ed.) *Hyman's Comparative Vertebrate Anatomy* (Univ. Chicago Press, Chicago, 1979).
- Crow, G. L. *et al. Copeia* **1990**, 226–229 (1990).
- Budker, P. *The Life of Sharks* (Weidenfeld & Nicolson, London, 1971).
- Beintema, A. *J. Nature* **352**, 480–481 (1991).

Semaphorin 3A is a chemoattractant for cortical apical dendrites

Franck Polleux, Theresa Morrow & Anirvan Ghosh

Department of Neuroscience, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205-2185, USA

The apical dendrites of pyramidal neurons integrate inputs from various cortical layers and are central to information processing. Here we show that the growth of apical dendrites towards the pial surface is regulated by a diffusible chemoattractant present at high levels near the marginal zone. A major component of this signal is semaphorin 3A (Sema3A), which was previously characterized as a chemorepellant for cortical axons. Soluble guanylate cyclase is asymmetrically localized to the developing apical dendrite, and is required for the chemoattractive effect of Sema3A. Thus the asymmetric localization of soluble guanylate cyclase confers distinct Sema3A responses to axons and dendrites. These observations reveal a mechanism by which a single chemotropic signal can pattern both axons and dendrites during development.

Pyramidal neurons convey information from the cerebral cortex to various cortical and subcortical targets^{1,2}. These neurons are characterized by an apical dendrite that extends towards the pial surface and integrates information from superficial cortical layers³. Despite the functional importance of apical dendrites in cortical information processing, the mechanisms that regulate their orientation are not known. One possibility is that correct orientation involves the

directed growth of the dendritic process towards the pial surface. Extracellular signals can influence dendritic growth and branching^{4–6}, so the directed dendritic growth might also be regulated by extracellular signals. We carried out a series of experiments to determine whether the directed growth of apical dendrites was regulated by signals in the extracellular environment. We were particularly interested in exploring the possibility that chemotropic

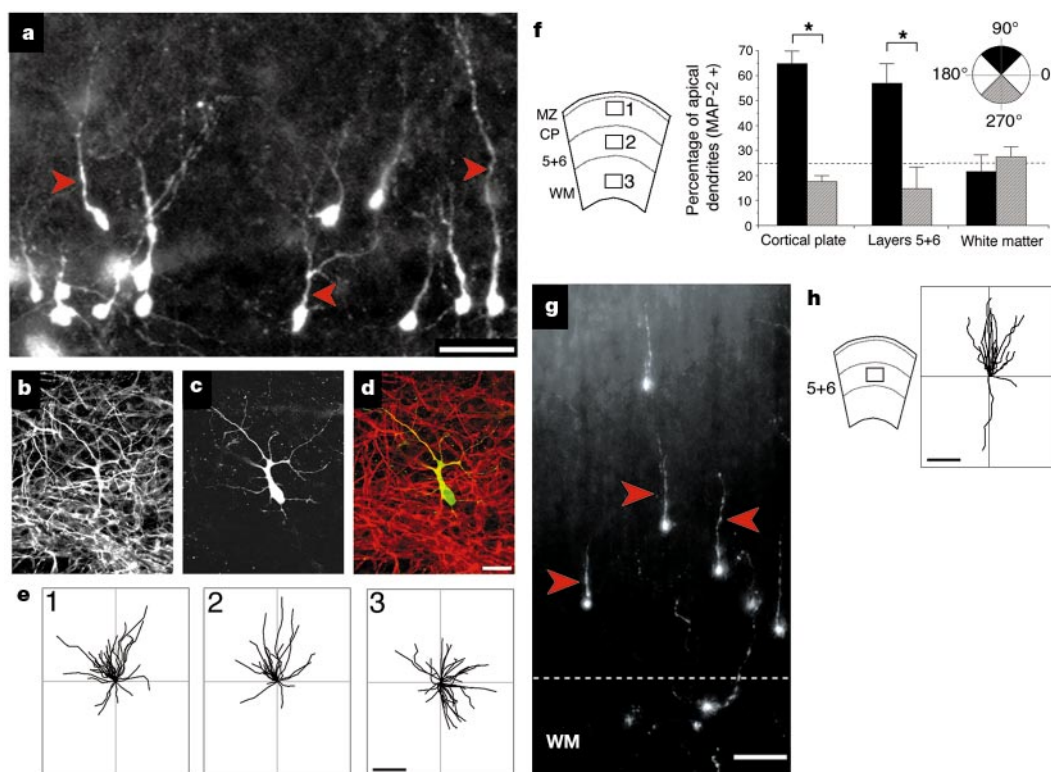


Figure 1 An extracellular signal present in the cerebral wall directs apical dendrite outgrowth towards the marginal zone. **a**, Examples of live E15 GFP neurons growing over a P3 cortical slice, photographed at five days *in vitro*. The apical dendrites (arrowheads) of GFP(+) pyramidal neurons are directed towards the marginal zone of the slice (top). **b–d**, Double immunofluorescence directed against MAP-2 (red channel, **b**) and GFP (green channel, **c**) indicates that the morphologically identified apical dendrites are immunopositive for the dendritic protein, MAP-2 (yellow on the merged image, **d**). **e**, Apical dendrite orientation plots of GFP neurons plated over different regions of a representative cortical slice (numbers refer to the inset in **f**: 1, cortical plate (CP); 2, layers

5 and 6; 3, white matter (WM)). **f**, Apical dendrite orientation histogram of E15 GFP neurons cultured on P2 cortical slices for five days. In this and other histograms, the horizontal dashed line indicates expected distribution for random orientation. Asterisk, $P < 0.001$ according to a χ^2 test. **g**, E15 GFP neurons transplanted into rat cortex at P1 differentiate into pyramidal neurons with an apical dendrite directed towards the pial surface (arrowheads). **h**, Apical dendrite orientation plot of E15 GFP neurons transplanted into rat P1 cortex, examined at P5. Most apical dendrites are directed towards the pial surface (78.6%; $n = 28$). Scale bars: **a**, 80 μm ; **b–d**, 25 μm ; **e** (1–3) 100 μm ; **g**, 60 μm ; **h**, 100 μm .

signals, such as those described for axon guidance^{7,8}, might be involved in the patterning of dendrites.

Regulation of dendrite orientation by a chemoattractant

To study the effect of the cortical environment on dendritic patterning, we modified a previous approach used to evaluate the influence of extracellular signals on axon guidance in the cortex⁹. We dissociated embryonic day 15 (E15) cortical neurons from green fluorescent protein (GFP)-expressing transgenic mice¹⁰ and plated them onto neonatal cortical slices in culture. After five days *in vitro*, we visualized the morphology of plated neurons using anti-GFP immunofluorescence. Morphological analysis indicated that about 70% of GFP-positive neurons differentiated into pyramidal neurons characterized by one major apical dendrite and several basal dendrites (Fig. 1a). The dendritic nature of the apical processes was confirmed by double immunofluorescence against GFP and the dendritic marker microtubule associated protein-2 (MAP-2; Fig. 1b–d). Analysis of apical dendrite orientation of GFP-expressing pyramidal neurons (GFP neurons) at various locations across the cerebral wall revealed that around 65% of the apical dendrites of the GFP neurons growing over the cortical plate (CP; Fig. 1e, left, f) and over layers 5 + 6 (Fig. 1e, middle, f) were oriented towards the pial surface, whereas only about 15% were directed towards the white matter. In contrast, the apical dendrites of GFP neurons growing over the white matter were randomly oriented (Fig. 1e, right, f). Thus an extracellular signal present in the developing cerebral wall is able to orient apical dendrite outgrowth towards the pial surface.

To confirm that such dendrite orientation signals were also present *in vivo*, we transplanted E15 cortical GFP neurons into rat cortex on the day of birth, and analysed the morphology of transplanted cells after five days. Many of the transplanted cells in the cortex differentiated into pyramidal neurons (Fig. 1g) and most of them (78.6%; $n = 28$) extended apical dendrites oriented towards the pial surface (Fig. 1h).

To determine whether the dendrite orientation signal was diffusible, we performed a slice co-culture experiment in which the white matter of one cortical slice was placed next to the marginal zone of a second slice for 2–3 h before E15 GFP neurons were plated over the slices (Fig. 2a–d). Normally apical dendrites of neurons plated over the white matter are oriented randomly (Figs 1e, right, 2c, inset 2). However, when the marginal zone of a second slice was placed next to the white matter of the first slice (Fig. 2c), apical dendrites growing on the white matter of the first slice were directed towards the marginal zone of the second slice (Fig. 2c, inset 1, d). The ability of a slice to influence dendrite orientation in an adjacent slice indicates that the dendrite orientation signal is diffusible. The fact that apical dendrites on either side of the slice culture interface are directed towards the marginal zone indicates that the apical dendrites are attracted by a signal present at high levels near the marginal zone.

Using a similar co-culture configuration we have shown that a diffusible chemorepellant directs cortical axons away from the marginal zone⁹. As the axon typically emerges before the apical dendrite, one interpretation of the co-culture experiment is that the factor released by the marginal zone instructively orients axon outgrowth and that the apical dendrites grow 'by default' in an opposite direction. To test whether the cortical signal could exert its effect on apical dendrite orientation independently of the initial axon trajectory, we performed a modified co-culture experiment. We grew GFP neurons on a cortical slice for one day before a second slice was placed adjacent to the white matter of the first slice (Fig. 2e). As cortical neurons growing over the white matter extend randomly oriented axons within a few hours after plating⁹, this experiment would reveal the influence of a second slice on dendritic orientation after initial axon extension. In this experiment the apical dendrites of GFP neurons growing over the white matter

of the first slice were predominantly oriented towards the marginal zone of the second slice (Fig. 2e, inset 2, f). Thus the cortical signal can act as an attractant for apical dendrites independently of initial axon trajectory.

To determine whether the growing apical dendrite can respond to changes in the distribution of extracellular guidance cues, we analysed apical dendrites that grew from one slice onto a second

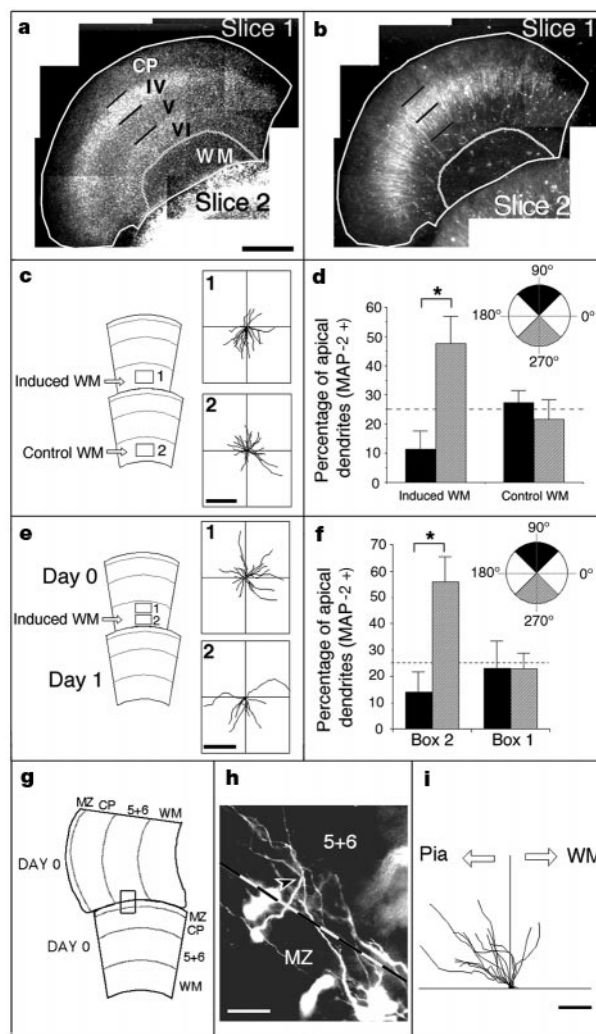


Figure 2 Apical dendrites are attracted by a diffusible signal near the marginal zone. **a, b**, The cytoarchitecture of cortical slices is preserved after six days in culture as revealed by nuclear staining using Hoechst 33258 (**a**) and dendritic MAP-2 labelling (**b**). **c**, Apical dendrite orientation plot of E15 GFP neurons cultured for four days over the induced white matter (inset 1) or the control white matter (inset 2) of P3 slices co-cultured as shown. **d**, Apical dendrite orientation histograms for the experiment shown in **c**. **e**, Apical dendrite orientations of GFP neurons cultured over P3 cortical slice in a delayed co-culture experiment where E15 GFP neurons were plated over a P3 slice on the first day (Day 0). Twenty-four hours after plating, a second cortical slice was placed next to the first slice (Day 1). Insets 1 and 2 illustrate the apical dendrite orientation of GFP neurons growing distant from (300–600 μm , inset 1) or close to (0–300 μm , inset 2) the slice–slice interface. **f**, Apical dendrite orientation histogram for the experiment depicted in **e**. **g–i**, Turning of apical dendrites towards the pial surface of an orthogonal slice. P3 slices were co-cultured for five days as shown in **g**, with the marginal zone of one slice in direct contact with the lateral aspect of a second slice. The box indicates the position of the GFP neurons shown in **h** whose apical dendrites turn towards the pial surface of the second slice (arrowhead). **i**, Trajectories of apical dendrites ($n = 15$) growing from the marginal zone of one slice into layers 5 and 6 or the cortical plate of a second slice positioned as depicted in **g**. Arrows indicate the direction to the pial surface and the white matter of the second slice. Scale bars: **a, b**, 600 μm ; **c**, 200 μm ; **e**, 200 μm ; **h**, 25 μm ; **i**, 40 μm .

slice positioned orthogonally (Fig. 2g). The majority (87%; $n = 15$) of apical dendrites that crossed from one slice to the orthogonal slice turned towards the pial surface of the second slice (Fig. 2h, i; average turning angle, $50.3^\circ \pm 9.2^\circ$). These observations indicate that growth of apical dendrites towards the pial surface is a chemotropic response.

Sema3A is a chemoattractant for apical dendrites

Cortical axons are directed towards the white matter by the diffusible chemorepellant Sema3A⁹, which is expressed in the cortical plate during development^{12–14,21} (see Supplementary Information). In our analysis of the SEMA3A null mice¹⁶ we noticed that many of the cortical pyramidal neurons had abnormal morphologies⁹, raising the possibility that Sema3A may be involved in regulating apical dendrite orientation. Therefore, we labelled cortical projection neurons in wild-type and SEMA3A null mice by injecting fluorescent carbocyanine dyes (DiI or DiA) into the white matter of cortical slices (Fig. 3a inset). In contrast to wild-type neurons, many of the neurons in SEMA3A null mice had identifiable apical dendrites that were not directed towards the pial surface (Fig. 3b–e). Other aspects of dendritic growth and branching were not notably affected in these mice. Thus Sema3A function appears to be necessary for correct orientation of apical dendrites.

If apical dendrite orientation is regulated by a gradient of Sema3A in the cortex, then abolishing that gradient by providing excess Sema3A should disrupt apical dendrite orientation. We cultured E15 GFP neurons on cortical slices for four days in the presence of exogenous recombinant alkaline phosphatase-tagged Sema3A

(Sema3A-AP). Whereas incubation of the slices with supernatant from control-transfected 293T cells did not affect the orientation of apical dendrites (Fig. 3f inset 1, g), incubation with supernatant from Sema3A-AP-transfected 293T cells completely disrupted apical dendrite orientation (Fig. 3f inset 2, g).

To determine whether a localized source of Sema3A could direct apical dendrite outgrowth, we placed control or Sema3A-expressing 293T cell aggregates next to the white matter of cortical slices over which we plated E15 GFP neurons. The GFP neurons extended apical dendrites towards the Sema3A-expressing aggregates (Fig. 3h inset 2, i), but were unaffected by control 293T cell aggregates (Fig. 3h inset 1, i). Thus, a localized source of Sema3A is sufficient to attract apical dendrites. Importantly, this effect of Sema3A on apical dendrite orientation is opposite to the chemorepulsive effect of Sema3A on cortical axons⁹.

The repulsive effect of Sema3A on axons is mediated by the high-affinity semaphorin receptor Neuropilin-1 (refs 9, 11, 15). To determine whether the directed growth of apical dendrites towards the pial surface was also mediated by Neuropilin-1, we examined the subcellular distribution of Neuropilin-1 using an affinity-purified antibody^{9,11}. Neuropilin-1 is present on both axons and dendrites of cortical neurons (Fig. 4a–c). To examine the role of Neuropilin-1 in dendrite orientation we incubated GFP slice overlay cultures with an affinity-purified function-blocking antibody to Neuropilin-1 (anti-Np1)^{9,11}. In control cultures about 70% of the dendrites of GFP neurons were oriented towards the pia, and less than 10% were directed towards the white matter (Fig. 4d–h). In anti-Np1-treated slices, only about 45% of the dendrites were directed towards the

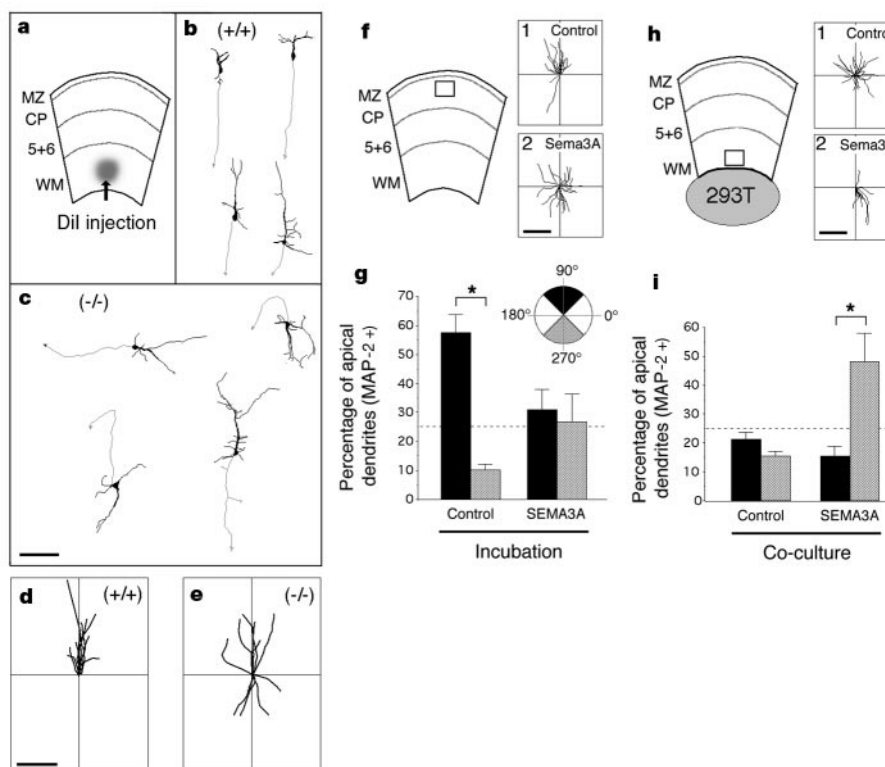


Figure 3 Sema3A is necessary and sufficient for appropriate apical dendrite orientation. **a**, Location of carbocyanine dye injection into the white matter of fixed cortical slices from E20 mice to label projection neurons retrogradely (reconstructed in **b** and **c**). **b**, **c**, Examples of cortical neuron morphologies in the cortical plate (two upper neurons) or in layers 5 and 6 (two bottom neurons) of wild-type (+/+) (**b**) or SEMA3A null (-/-) (**c**) mice. The pia is at the top. **d**, **e**, Apical dendrite orientation plot of cortical projection neurons in E20 wild-type (+/+) (**d**) or SEMA3A null (-/-) mice (**e**). **f**, **g**, Incubation with excess of recombinant Sema3A-AP disrupts the orientation of apical dendrite outgrowth in the cortical plate. **f**, Apical dendrite orientation plot of E15 GFP neurons cultured for four

days over the cortical plate of a P3 slice in the presence of control supernatant or supernatant from Sema3A-AP expressing cells (Sema3A-AP concentration 8 nM). **g**, Apical dendrite orientation histogram for the experiment shown in **f**. **h**, **i**, A localized source of Sema3A is sufficient to attract apical dendrites. **h**, Apical dendrite orientation plot of E15 GFP neurons cultured for four days over the white matter of a P3 slice next to an aggregate of 293T cells transfected with a control vector (inset 1) or a plasmid expressing Sema3A (inset 2). **i**, Apical dendrite orientation histogram for the experiment depicted in **h**. Scale bars: **b**, **c**, 100 μ m; **d**, **e**, 120 μ m; **f**, **h**, 200 μ m.

pia, and over 20% were directed towards the white matter (Fig. 4h). The marked reduction in the bias of apical dendrites to grow towards the pial surface indicates that the response of the dendrites to the cortical guidance signal is mediated at least in part by Neuropilin-1. The residual bias in the growth of apical dendrites towards the pial surface might simply reflect incomplete blocking over a four-day culture period, or might reflect a component of the dendritic response that is Neuropilin-1-independent. In either case, the ability of anti-Np1 to attenuate the dendrite orientation response is consistent with the experiments that show that this process is regulated by *Sema3A*.

Role of soluble guanylate cyclase in dendrite patterning

How could *Sema3A* act as a chemorepellant for the axon and a chemoattractant for the apical dendrite of the same neuron? One possibility is that axons and dendrites express different *Sema3A* receptors, but the Neuropilin-1 localization experiments indicate that at least one critical component of the receptor complex is present on both axons and dendrites. Another possibility is that a signal transduction component necessary for *Sema3A* signalling may be asymmetrically localized to the axon or apical dendrite. In that case, the signal transduction component must specify the polarity of the cellular response to *Sema3A*. Asymmetric localization of soluble guanylate cyclase (SGC) could in principle serve such a function, as the chemorepulsive response of *Xenopus* spinal cord axons to *Sema3A* can be converted into chemoattraction by

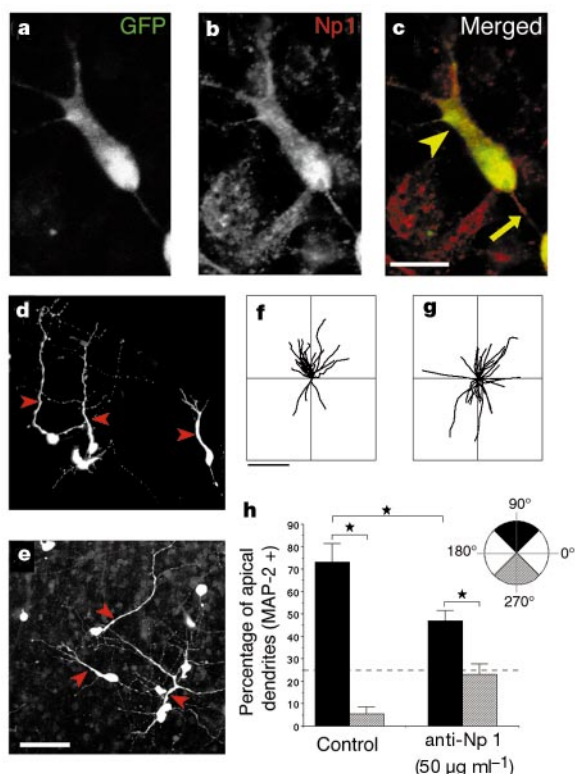


Figure 4 The *Sema3A* receptor Neuropilin-1 is present on apical dendrites, and is required for correct apical dendrite orientation. **a–c**, Double immunofluorescence against GFP (**a**) and Neuropilin-1 (Np1) (**b**) on a GFP neuron cultured for one day over the cortical plate of a P3 cortical slice shows that Neuropilin-1 is present on both the axon (arrow in **c**) and the apical dendrite (arrowhead in **c**) of developing cortical neurons. In the merged image (**c**), GFP is in green and Neuropilin-1 is in red. **d, e**, Examples of GFP neurons growing over the cortical plate of a P3 cortical slice for four days in the absence (**d**) or presence (**e**) of affinity-purified anti-Np1. **f, g**, Apical dendrite orientation plots of E15 GFP neurons growing over the cortical plate of postnatal cortical slices in the absence (**f**) or presence (**g**) of anti-Np1. **h**, Apical dendrite orientation histogram for the experiment described in **f, g**. Scale bars: **a–c**, 15 μ m; **d, e**, 50 μ m; **f, g**, 200 μ m.

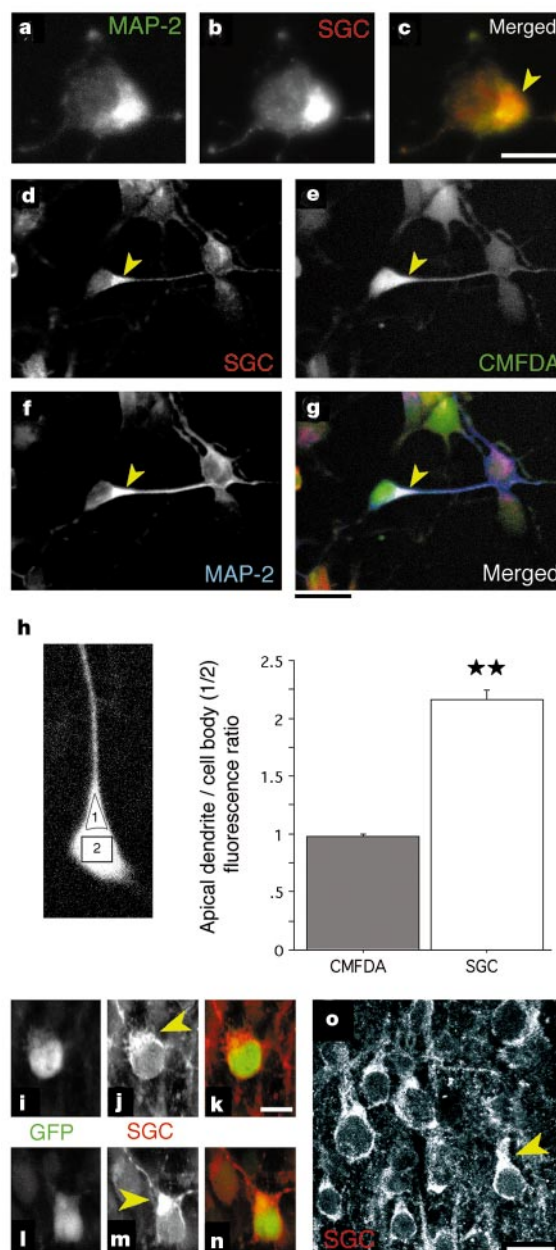


Figure 5 Soluble guanylate cyclase is asymmetrically localized to the apical dendrite in developing pyramidal neurons. **a–c**, MAP-2 (**a**) and SGC (**b**) immunofluorescence in cortical neurons 3 h after plating. The merged image (**c**) shows that SGC (red) co-localizes with MAP-2 (green) in the region of the cell where the apical dendrite is generated (arrowhead in **c**). **d–g**, Localization of SGC in pyramidal neurons in culture 48 h after plating. Cells were labelled with antibodies to SGC (**d**) and MAP-2 (**f**), and the distribution of the cytoplasm was visualized by loading with CMFDA (**e**). In the merged image (**g**), SGC is in red, MAP-2 is in blue and CMFDA is in green. **h**, Quantification of relative SGC immunofluorescence in apical dendrites and cell bodies of pyramidal neurons. Average fluorescence signals for CMFDA and SGC were quantified in regions 1 and 2, which represent the apical dendrite and cell body, respectively, and plotted as a fluorescence ratio. Double asterisk, $P < 0.01$ (Wilcoxon rank non-parametric test for paired values). **i–n**, SGC localization in two morphologically undifferentiated E15 GFP neurons (**i–k** and **l–n**) that were cultured for 12 h over the cortical plate of a P3 slice. SGC (arrows in **j** and **m**) is asymmetrically localized in a cap oriented towards the pial surface (arrows) before an apical dendrite is clearly apparent. **o**, Subcellular distribution of SGC in layer 5 neurons of P3 cortex showing its enrichment in the apical dendrites of large pyramidal neurons. Scale bars: **a–c**, 10 μ m; **d–g**, 30 μ m; **i–n**, 10 μ m; **o**, 15 μ m.

artificially elevating intracellular cyclic GMP (cGMP) levels¹⁷.

To explore the possibility that SGC might be asymmetrically localized in cortical neurons, we used immunofluorescence to determine the subcellular distribution of SGC. In dissociated neurons beginning to undergo morphological differentiation in culture, SGC was localized to the pole of the cell where the apical dendrite is generated (Fig. 5a–c). Strikingly, a cap of SGC co-localized with MAP-2 was apparent even before the neurons had extended a morphologically distinguishable apical dendrite (Fig. 5a–c), indicating that SGC might be important in the early patterning of the apical dendrite. To confirm that the apparent asymmetric localization of SGC was not simply due to increased cytoplasm at one pole of the cell, we triple-labelled cortical neurons with anti-SGC, anti-MAP-2 and the vital dye CMFDA (a cytoplasmic marker). Whereas CMFDA immunofluorescence was uniformly distributed over the cell body and apical dendrite (Fig. 5e, g, h), SGC and MAP-2 immunofluorescence was localized to the apical dendrite (Fig. 5d, f–h). SGC immunofluorescence in undifferentiated GFP neurons 12 h after being plated over the cortical plate of a postnatal day 2 (P2) slice showed asymmetric caps of SGC oriented towards the pial surface before the formation

of apical dendrites (Fig. 5i–n). We also found that SGC was localized to the base and proximal segment of apical dendrites of layer 5 pyramidal neurons in P3 cortical slices, indicating that the asymmetric subcellular distribution of SGC is characteristic of cortical neurons *in vivo* (Fig. 5o).

The localization of SGC to the developing apical dendrite together with evidence implicating cGMP in Sema3A signalling suggested that SGC might be important for the oriented growth of apical dendrites towards the pial surface. Therefore, we performed slice overlay assays in the presence of ODQ (1H-[1,2,4]oxadiazolo[4,3-a]quinoxalin-1-one), a specific inhibitor of SGC¹⁸. Whereas apical dendrite orientation was unaffected in vehicle (DMSO)-treated slices (Fig. 6a, f, k), ODQ treatment (1 μ M) completely disrupted oriented dendritic outgrowth (Fig. 6b, g, k). The differentiation of the apical dendrite, however, was not affected by ODQ treatment (Fig. 6b), indicating that SGC is not required for dendritogenesis but is specifically required for oriented dendritic outgrowth. To determine whether the effects of SGC were mediated by the downstream effector cGMP-dependent protein kinase (PKG), we treated slice overlay cultures with Rp-8-pCPT-cGMPs (8-(4-Chlorophenylthio)-guanosine-3'-5' cyclic monophosphoro-

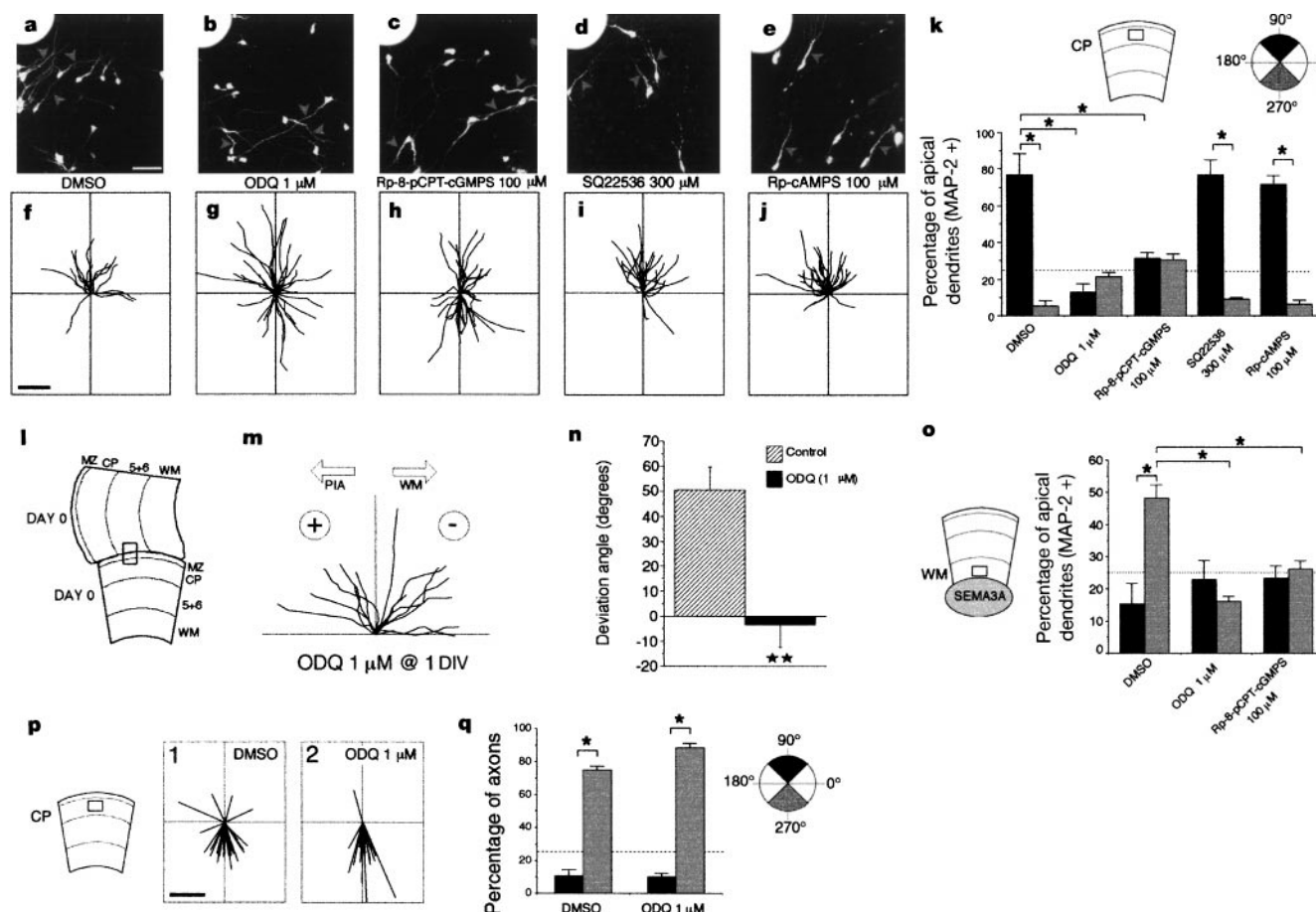


Figure 6 Inhibition of soluble guanylate cyclase and protein kinase G disrupts apical dendrite orientation. **a–j**, Photomicrographs and apical dendrite orientation plots of E15 GFP neurons cultured over the cortical plate of P2 cortical slices for three days in the presence of DMSO (**a, f**; control vehicle at a 1:1,000 dilution), 1 μ M ODQ, an inhibitor of SGC (**b, g**), 100 μ M Rp-8-pCPT-cGMPs, an inhibitor of all protein kinase G isoforms (**c, h**), 300 μ M SQ22536, an inhibitor of all isoforms of adenylate cyclase (**d, i**), 100 μ M Rp-cAMPS, an inhibitor of protein kinase A (**e, j**). ODQ- and Rp-8-pCPT-cGMPs-treated cells extend apical dendrites (arrowheads), but they are randomly oriented. **k**, Apical dendrite orientation histogram for the experiment shown in **a–j**. **l–n**, Effect of ODQ on the ability of apical dendrites to reorient in response to an orthogonally positioned slice (shown in **l**). **m**, Trajectories of apical dendrites of GFP neurons plated in the lower slice (**l**) that cross

into the orthogonal slice exposed to 1 μ M ODQ after one day *in vitro* (DIV). **n**, Comparison of average turning angles of apical dendrites (measured at 5 DIV) in the absence or presence of ODQ. Positive deviation angles represent deflections towards the pial surface of the second slice. Double asterisk, $P < 0.01$ (Mann-Whitney non-parametric test). **o**, Apical dendrite orientation histogram of GFP neurons cultured for four days over the white matter of P2 slices placed next to an aggregate of Sema3A-expressing 293T cells in the presence of indicated pharmacological agents. **p, q**, Inhibition of SGC activity by ODQ does not affect directed axon outgrowth towards the white matter. **p**, Axon orientation plots of cortical neurons cultured on P2 slices for 3 h in the presence of control vehicle (DMSO, 1:1,000; inset 1) or 1 μ M ODQ (inset 2). **q**, Axon orientation histogram for the experiment described in **p**. Scale bars: **a–e**, 60 μ m; **m**, 150 μ m; **p**, 40 μ m.

thioate, Rp isomer; 100 μ M), a specific inhibitor of all three PKG isoforms (I α , I β and II)¹⁹. As with SGC, inhibition of PKG disrupted the oriented growth of apical dendrites towards the pial surface (Fig. 6c, h, k). Apical dendrite orientation appears to involve the cGMP-dependent pathway specifically, as incubation with a general adenylate cyclase inhibitor, SQ22586 (9-(tetrahydro-2'-furyl)-adenine; 300 μ M)²⁰ or with the protein kinase A inhibitor Rp-cAMPS (adenosine 3'-5'-cyclic monophosphorothioate, Rp-isomer; 100 μ M) did not affect apical dendrite orientation (Fig. 6d, i, k and Fig. 6e, j, k, respectively). In further support of a role for SGC in dendritic guidance, we found that the ability of apical dendrites to turn towards the pial surface of an orthogonal cortical slice was completely disrupted in ODQ-treated cultures (Fig. 6l–n).

To determine whether the chemoattractive effect of *Sema3A* is also regulated by SGC- and PKG-dependent mechanisms, we performed slice-overlay assays with *Sema3A*-expressing 293T cells placed next to the white matter in the presence of various pharmacological inhibitors. Control treatments did not affect the growth of apical dendrites towards the *Sema3A*-expressing cells, but inhibition of SGC (1 μ M ODQ) or PKG (100 μ M Rp-8-pCPT-cGMPS) completely disrupted the oriented outgrowth of the apical dendrites towards the *Sema3A* source (Fig. 6o). Thus the chemoattractive effect of *Sema3A* on apical dendrites requires SGC and PKG function.

Finally, we examined the effects of inhibiting SGC on directed axon outgrowth to determine whether cGMP signalling is specifically involved in apical dendrite orientation. ODQ treatment did not affect the directed growth of cortical axons towards the white matter (Fig. 6p, q). Thus, inhibition of SGC does not have a general disruptive effect on all chemotropic responses, and specifically affects the oriented growth of apical dendrites.

Discussion

These experiments provide the first evidence that different compartments of a neuron can respond in opposite ways to the same chemotropic signal: apical dendrites of pyramidal neurons grow towards a source of *Sema3A*, whereas their axons are repelled by the same ligand. *Sema3A* is expressed in the cortex, but the pattern of expression changes during development, so it may be involved in regulating several developmental events. At the onset of cortical neurogenesis, *Sema3A* is expressed primarily within the ventricular zone, which is consistent with the proposal that *Sema3A* may prevent early cortical efferents (such as those from subplate neurons) from entering the ventricular zone^{21,22}. At late embryonic and early postnatal stages, *Sema3A* is expressed within the cortical plate^{12–14,21} (see Supplementary Information). Although *Sema3A* appears to be expressed in all cortical layers at these stages, the higher density of neurons in the superficial cortical plate may allow a gradient of *Sema3A* to be established. We propose that a polarized cellular response to *Sema3A* causes cortical neurons to extend efferent axons towards the white matter and apical dendrites towards the pial surface.

The differential response of axons and dendrites to *Sema3A* appears to be mediated by the asymmetric localization of SGC to the developing apical dendrite. Our results indicate that localization of SGC to the dendrite confers upon it a chemoattractive response, which is consistent with the observation that elevation of cGMP levels can allow *Sema3A* to act as a chemoattractant¹⁷. The generation of a polarized guidance response by asymmetric localization of an intracellular signalling component is similar to a mechanism that has been proposed for chemotaxis in *Dictyostelium* and mammalian leukocytes^{23,24}. In these cells, the receptor for a chemotactic molecule is distributed uniformly over the cell surface and a chemotactic response is triggered by the asymmetric distribution of one component of the signal transduction machinery. Asymmetric localization of signal transduction components may therefore be a general

mechanism for generating polarized cellular responses^{25,26}.

Little is known about the mechanism by which cGMP signalling might modulate responses to *Sema3A*, but our pharmacological experiments indicate that PKG is likely to be an important downstream effector in this process. Also inhibition of SGC or PKG disrupts oriented dendritic growth but does not cause the dendrites to be repelled by *Sema3A*. Thus, SGC does not simply act as a polarity switch, and repulsion of axons by *Sema3A* must involve mechanisms that are not functional in dendrites. The polarity of chemotropic responses to Netrin-1 can be influenced by differences in Netrin-1 receptor complexes^{27,28}, indicating that differences in receptor complexes between axons and dendrites might also contribute to the differential responses to *Sema3A*. The details of the intracellular signalling mechanisms that generate diverse *Sema3A* responses remain to be understood. Our results provide compelling evidence that the patterning of dendrites can be regulated by chemotropic cues, and that the differential response of axons and dendrites to such cues can be an important determinant of neuronal morphology. □

Methods

Slice overlay assay

Early postnatal rat cortical slices (P1–P3) were cultured using an air-interface protocol⁹. Embryonic cortical neurons were isolated at E15 from timed pregnant transgenic mice expressing the enhanced green fluorescent protein (EGFP) under the control of a β -actin promoter¹⁰. After gentle trypsin-based dissociation of the cortex, GFP(+) cells were cultured on top of cortical slices for 4–6 days.

Immunofluorescence

Slices were fixed overnight with 4% paraformaldehyde and processed for immunofluorescence using a protocol where the three major steps (blocking of non-specific antigenic sites in 3% BSA plus 0.3% Triton X-100, incubation with primary antibody and incubation with fluorescent secondary antibodies) were performed overnight at 4 °C. The primary antibodies used were: rabbit polyclonal anti-GFP (1:3,000; Molecular Probes), mouse monoclonal anti-GFP (1:2,000, Molecular Probes), mouse monoclonal anti-MAP-2 (1:3,000, Sigma), rabbit polyclonal anti- β 1-subunit of soluble guanylate cyclase (1:2,000, Calbiochem), and a rabbit polyclonal, affinity-purified, anti-Neuropilin-1 (1:600, A. Kolodkin and D. Ginty). The fluorescent secondary antibodies were goat anti-rabbit or goat anti-mouse Oregon Green-conjugated IgG (1:800, Molecular Probes), goat anti-rabbit or goat anti-mouse Cy3-conjugated IgG and donkey anti-mouse Cy5-conjugated IgG (1:800, Jackson ImmunoResearch). The secondary antibody incubations were performed in the presence of 3% goat or normal donkey serum. Cell nuclei were stained for 2 h using Hoechst 33258 (1:2,000, Molecular Probes).

Scoring of apical dendrite orientation

Apical dendrites were imaged and scored using a Nikon TE300 Eclipse microscope equipped with 20 $\times$ and 40 $\times$ ELWD objectives. GFP(+) neurons were imaged and analysed using IPLab Spectrum 3.2 (Scanalytics). To generate apical dendrite orientation plots, the trajectories of individual apical dendrites were plotted with the pial surface at the top (90°) in each plot (for example, Fig. 1e). Apical dendrites were scored as being directed towards the pia (45°–135°) or towards the ventricle (225°–315°) based on their orientation. To generate apical dendrite orientation histograms, the orientation of MAP-2(+) apical dendrites of GFP neurons was scored for a minimum of 150 neurons from six to twelve slices for each experimental condition.

In vivo transplantation of E15 GFP(+) cells

10,000–20,000 freshly isolated E15 GFP(+) cortical cells were transplanted using a Hamilton syringe into the cortex of hypothermia-anaesthetized rats at P1. After five days survival, pups were perfused using 4% paraformaldehyde, sectioned at 100 μ m using a vibratome, and processed for immunofluorescence directed against MAP-2 and GFP.

SEMA3A knockout mice

A colony of *SEMA3A* knockout mice¹⁶ was maintained using heterozygous animal breeding. Experimental litters were perfused at E20 and treated blind with regard to the genotype of the embryos (analysed by PCR using tail DNA).

Tracing experiments

The morphology of cortical projection neurons was reconstructed using fluorescent carbocyanine dye (DiI or DiA, Molecular Probes) injected into the white matter of fixed coronal slices from E20 wild-type ($n = 11$) and *SEMA3A* null ($n = 8$) mice. After 2–3 weeks of diffusion, 80–100 μ m thick vibratome sections were studied using confocal microscopy.

Microscopy

Reconstruction of DiI- and DiA-labelled neurons (Fig. 3), and immunofluorescent localization of proteins (Figs 4, 5), was carried out using an LSM 510 Zeiss confocal microscope equipped with Helium-Neon and Argon lasers.

CMFDA labelling

Dissociated E15 cortical cells in suspension were centrifuged once and incubated for 30 min at 37 °C in culture medium containing 10 µM CellTracker Green CMFDA (5-chloromethylfluorescein diacetate, Molecular Probes). Cells were then incubated for 30 min in CMFDA-free culture medium at 37 °C. During this second incubation CMFDA undergoes a glutathione S-transferase-mediated reaction to produce membrane-impermeant glutathione-fluorescent dye adducts that can be detected by direct fluorescence using a 488-nm excitation line.

Production of 293T cell aggregates

293T cells (ATCC) were transfected using Lipofectamine+ (Gibco) according to a standard protocol²⁹. Sema3A-expressing 293T cells were produced by co-transfection of a plasmid expressing Sema3A protein tagged with alkaline phosphatase⁹ and an EGFP-expressing plasmid (Clontech) for a better visualization of the aggregate-slice boundary. Control-transfected 293T cells were transfected only with a plasmid expressing EGFP.

Production of recombinant Sema3A-AP protein

293T cells cultured in serum-free medium (Nephrogen-Celox) were transfected as described above, and after two days supernatant of control or Sema3A-AP-transfected 293T cells was collected and concentrated using a centrifugation-based dialysis device (Centriprep, Millipore). The final concentration of Sema3A-AP was determined using an alkaline phosphatase enzymatic assay³⁰.

Scoring of axon orientation

Cortical neurons were isolated from E18 rat cortex, dissociated, labelled with DiI and plated over P2 rat cortical slices⁹. They were co-cultured in the presence of control vehicle (DMSO, 1:1,000) or 1 µM ODQ and then axon orientation was determined⁹.

Received 25 October 1999; accepted 16 February 2000.

1. Miller, M. W. in *Cerebral Cortex* (eds Peters, A. & Jones, E. G.) 133–175 (Plenum, New York, London, 1988).
2. Gilbert, C. D. Microcircuitry of the visual cortex. *Annu. Rev. Neurosci.* **6**, 217–247 (1983).
3. Larkum, M. E., Zhu, J. J. & Sakmann, B. A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature* **398**, 338–341 (1999).
4. Fishell, G. Striatal precursors adopt cortical identities in response to local cues. *Development* **121**, 803–812 (1995).
5. McAllister, A. K., Lo, D. C. & Katz, L. C. Neurotrophins regulate dendritic growth in developing visual cortex. *Neuron* **15**, 791–803 (1995).
6. Wu, G. & Cline, H. Stabilization of dendritic arbor structure in vivo by CaMKII. *Science* **279**, 222–226 (1998).
7. Tessier-Lavigne, M. & Goodman, C. S. The molecular biology of axon guidance. *Science* **274**, 1123–1133 (1996).
8. Kolodkin, A. L. Semaphorin-mediated neuronal growth cone guidance. *Prog. Brain Res.* **117**, 115–132 (1998).
9. Polleux, F., Giger, R. J., Ginty, D. D., Kolodkin, A. L. & Ghosh, A. Patterning of cortical efferent projections by semaphorin-neuropilin interactions. *Science* **282**, 1904–1906 (1998).
10. Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* **407**, 313–319 (1997).

11. Kolodkin, A. L. *et al.* Neuropilin is a semaphorin III receptor. *Cell* **90**, 753–762 (1997).
12. Catalano, S. M., Messersmith, E. K., Goodman, C. S., Shatz, C. J. & Chédotal, A. Many major CNS axon projections develop normally in the absence of semaphorin III. *Mol. Cell. Neurosci.* **11**, 173–182 (1998).
13. Chédotal, A. *et al.* Semaphorins III and IV repel hippocampal axons via two distinct receptors. *Development* **125**, 4313–4323 (1998).
14. Skaliara, I., Singer, W., Betz, H. & Puschel, A. W. Differential patterns of semaphorin expression in the developing rat brain. *Eur. J. Neurosci.* **10**, 1215–1229 (1998).
15. He, Z. & Tessier-Lavigne, M. Neuropilin is a receptor for the axonal chemorepellent Semaphorin III. *Cell* **90**, 739–751 (1997).
16. Behar, O., Golden, J. A., Mashimo, H., Schoen, F. J. & Fishman, M. C. Semaphorin III is needed for normal patterning and growth of nerves, bones and heart. *Nature* **383**, 525–528 (1996).
17. Song, H.-J. *et al.* Conversion of neuronal growth cone responses from repulsion to attraction by cyclic nucleotides. *Science* **281**, 1515–1518 (1998).
18. Garthwaite, J. *et al.* Potent and selective inhibition of nitric oxide-sensitive guanylyl cyclase by 1H-[1,2,4]oxadiazolo[4,3-a]quinoxalin-1-one. *Mol. Pharmacol.* **48**, 184–188 (1995).
19. Butt, E., Eigenthaler, M. & Genieser, H. G. (Rp)-8-pCPT-cGMPS, a novel cGMP-dependent protein kinase inhibitor. *Eur. J. Pharmacol.* **269**, 265–268 (1994).
20. Harris, D. N., Asaad, M. M., Phillips, M. B., Goldenberg, H. J. & Antonaccio, M. J. Inhibition of adenylate cyclase in human blood platelets by 9-substituted adenine derivatives. *J. Cyclic Nucleotide Res.* **5**, 125–134 (1979).
21. Giger, R. J., Wolfer, D. P., De Wit, G. M. & Verhaagen, J. Anatomy of rat semaphorin III/collapsin-1 mRNA expression and relationship to developing nerve tracts during neuroembryogenesis. *J. Comp. Neurol.* **375**, 378–392 (1996).
22. Bagnard, D., Lohrum, M., Uziel, D., Puschel, A. W. & Bolz, J. Semaphorins act as attractive and repulsive guidance signals during the development of cortical projections. *Development* **125**, 5043–5053 (1998).
23. Parent, C. A., Blacklock, B. J., Froehlich, W. M., Murphy, D. B. & Devreotes, P. N. G protein signaling events are activated at the leading edge of chemotactic cells. *Cell* **95**, 81–91 (1998).
24. Parent, C. A. & Devreotes, P. N. A cell's sense of direction. *Science* **284**, 765–770 (1999).
25. Baas, P. W. Microtubules and neuronal polarity: lessons from mitosis. *Neuron* **22**, 23–31 (1999).
26. Robinson, K. R. & Miller, B. J. The coupling of cyclic GMP and photopolarization of *Pelvetia* zygotes. *Dev. Biol.* **187**, 125–130 (1997).
27. Hong, K. *et al.* A ligand-gated association between cytoplasmic domains of UNC5 and DCC family receptors converts netrin-induced growth cone attraction to repulsion. *Cell* **97**, 927–941 (1999).
28. Bashaw, G. J. & Goodman, C. S. Chimeric axon guidance receptors: the cytoplasmic domains of slit and netrin receptors specify attraction versus repulsion. *Cell* **97**, 917–926 (1999).
29. Giger, R. J. *et al.* Neuropilin-2 is a receptor for semaphorin IV: insight into the structural basis of receptor function and specificity. *Neuron* **21**, 1079–1092 (1998).
30. Flanagan, J. G. & Leder, P. The kit ligand: a cell surface molecule altered in steel mutant fibroblasts. *Cell* **63**, 185–194 (1990).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature. It is also available on the authors' World-Wide Web site (<http://www.jhu.edu/ghosh>).

Acknowledgements

We thank A. Kolodkin and D. Ginty for the AP-Sema3A plasmid and the anti-Neuropilin-1 antibody, R. Giger for sharing data on SEMA3A expression in the cortex, M. Okabe for the GFP-expressing mice, O. Behar and M. Fishman for SEMA3A heterozygous mice, members of the Ghosh lab for discussions, and A. Kolodkin, D. Ginty, R. Giger and M. Tessier-Lavigne for comments on the manuscript. This work was supported by the NIH (A.G.) and the Pew Scholars Program (A.G.).

Correspondence and requests for materials should be addressed to A.G. (e-mail: aghosh@jhmi.edu).

Identification of comet Hyakutake's extremely long ion tail from magnetic field signatures

Geraint H. Jones*, André Balogh† & Timothy S. Horbury†

* Space and Atmospheric Physics Group, The Blackett Laboratory, Imperial College, London SW7 2BW, UK

† Astronomy Unit, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK

Observations of the varying orientations of comet tails led to the suggestion of the existence of the solar wind—a continuous outflow of ionized material from the Sun¹. It is now well established that gas from comets is ionized by several processes and joins the solar wind², forming an ion (plasma) tail that points away from the Sun. The plasma environments of three comets have been measured *in situ*, but only in the upstream direction or less than 8,000 km downstream of the nucleus. Here we report a fortuitous crossing by a spacecraft of the plasma tail of comet Hyakutake (C/1996 B2), at a distance of more than 3.8 astronomical units (550 million kilometres) from its nucleus. This surpasses the tail length of 2 AU determined for the Great March Comet of 1843 (C/1843 D1)³. Our measurements reveal that, at this distance, the tail of comet Hyakutake was a structured entity at least 7 million kilometres in diameter.

The Ulysses spacecraft (the first to explore the out-of-ecliptic solar wind) recorded anomalously low solar-wind proton number densities for several hours on 1 May 1996⁴. An unusual magnetic field structure was detected simultaneously by the spacecraft's magnetometer⁵. Analysis of the magnetic field data revealed that this structure displayed draping patterns consistent with those expected within a cometary plasma tail. Our search for the source comet led to analysis of the relative locations of Ulysses and the then-productive comet Hyakutake. On the day of the proton number-density drop, Ulysses was radially aligned with the position that Hyakutake had occupied approximately 8 days earlier. This

time difference was in the range of travel times expected for cometary plasma to traverse the space between the comet and spacecraft. The highly unusual alignment of the two objects coinciding with the 1 May event is strong evidence that Ulysses crossed Hyakutake's plasma tail (Fig. 1).

Hyakutake remained well away from the heliospheric current sheet during the period under study; no disconnection events⁶ were therefore expected to have occurred. The tail was thus likely to have been a single coherent structure from the nucleus to Ulysses, implying a tail length greater than 3.8 AU. The uniform, fast polar solar wind in which the comet was immersed at this time was also conducive to the preservation of the tail's integrity over exceptional distances. In other regimes, such as near-ecliptic alternating slow and fast streams, velocity shears would have made the recognition of the tail's magnetic field signature less likely.

Confirmation that the 1 May event was a tail crossing comes from more detailed analysis of the Ulysses data. Apart from the remarkable changes in the proton number densities, the solar wind did not stand out as being unusual. If viewed only in the context of the surrounding flow, the magnetic field structure would most probably not have been noticed. The magnetic field magnitude did not show a strong increase, and no shock wave was seen. It was only when attention was drawn to the structure, as a result of the behaviour⁴ of the solar-wind protons, that its uniqueness became apparent. All of the structure's unusual characteristics are consistent with a plasma tail crossing, and with that tail belonging to Hyakutake.

Deceleration of the solar wind occurs at comets, as a result of ion pick-up during momentum conservation. A velocity shear results⁷, giving entrained field lines hairpin-like configurations⁸ straddling an induced current sheet⁹. Evidence of velocity shear at Ulysses was a radial velocity decrease⁴ from ~ 750 to ~ 740 km s⁻¹; this small decrease is consistent with re-acceleration of plasma in the tail. The structure's field lines were clearly draped (Fig. 2). Near-radial solar wind flow would confine the tail to Hyakutake's orbital plane. The inferred current sheet orientations were indeed consistent with a tail axis parallel to this plane.

Using the comet's orbital velocity on 23 April, and a solar-wind velocity of 740 km s⁻¹, an expected angle from the anti-sunward (radial) direction of $\sim 38^\circ$ was calculated. The magnetic field vectors and discontinuities within the structure were aligned such that they

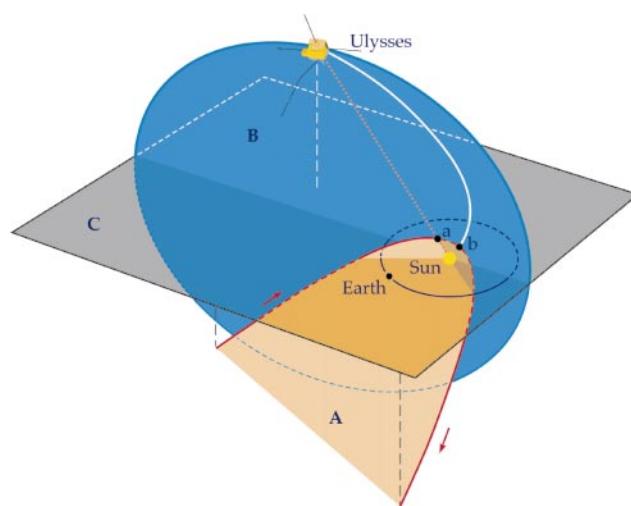


Figure 1 Relative positions of the Sun, comet Hyakutake, Ulysses and Earth on 1 May 1996. The orbital planes of the last three objects are labelled A, B and C, respectively. Ulysses was near Hyakutake's orbital plane, at 3.73 AU from the Sun ($1 \text{ AU} = 149.6 \times 10^6 \text{ km}$), at a solar ecliptic latitude of $\sim 45^\circ$ north, moving southwards at 6 arcmin d^{-1} . The comet had been at the same longitude as Ulysses' 1 May position around April 23.4 UT, and at Ulysses' 1 May latitude around April 24.0 UT (at heliocentric distances of around 0.35 AU; a). By 1 May, the comet was at perihelion, at 0.23 AU from the Sun (b). The

nucleus thus moved a considerable distance around its retrograde orbit (from a to b) while part of the tail travelled to Ulysses. The entire tail (shown in white) moved radially outwards which, combined with the comet's rapid motion, resulted in its strong curvature. Assuming a rapid acceleration to the solar-wind velocity, it approximated an archimedean spiral in the comet's orbital plane, and measured >3.8 AU in length. Assuming the region encountered by Ulysses originated at a during 23 April, its radial path of 3.39 AU (dotted line) took 7.4–8.0 days to traverse.

were consistent with a tail angle of between 63° and 70° . The observed angle was thus greater than expected: there are two possible reasons for this. The velocity shear between the slow-moving tail and the surrounding wind could have accentuated the angle, as is seen when a fast solar-wind stream impinges on a slower one. Alternatively, it could be argued that, as the tail section traversed by Ulysses was further from the Sun than some 'older' sections of the tail (which were carried outwards by slower wind), the tail angle could have been rotated in the sense seen. This latter explanation is unlikely, however, as only the tail section spanning the transition from slow to fast wind was likely to have been affected.

In reality, the comet's tail would take a finite time to accelerate up to the solar-wind speed⁷. Assuming the extreme case of a plasma starting from rest, and accelerating at a constant rate until reaching Ulysses, an acceleration rate of 1.1 m s^{-2} is inferred, which is consistent with values deduced from ground-based cometary tail observations ($0.15\text{--}3 \text{ m s}^{-2}$; ref. 10). As the centre of the tail was not crossed, the plasma was unlikely to have been initially stagnant, which suggests a lower acceleration rate. In all likelihood, the actual acceleration rate probably varied as the angle between the flow direction and the tail axis changed.

Taking 07:35 to 10:30 UT as representing the tail (that is, the times of the probable outermost tail magnetic signatures, broadly matching the limits of the proton number-density 'hole'), Ulysses' path across it measured $7.8 \times 10^6 \text{ km}$. With the tail's orientation known, and the fact that no current sheet was crossed, this gives a minimum tail diameter of $\sim 7 \times 10^6 \text{ km}$. However, the tail's cross-section may have been non-circular. Upstream solar-wind pressure may have reduced the tail's anti-solar extent, and the tail's expansion as it moved away from the Sun—in combination with rotation away from the radial direction—also suggests a non-circular tail profile.

Further evidence of the structure's source being Hyakutake comes from its size. Calculation of Hyakutake's tail diameter requires knowledge of the comet's local solar-wind parameters and its gas production rate. Few relevant ground-based observations were performed around 23 April. However, tentative extrapolation of

earlier production rates¹¹, and application of an empirical formula¹² to published visual magnitudes¹³, yields a production rate of $(1.0\text{--}2.2) \times 10^{30} \text{ molecules s}^{-1}$. When the solar-wind parameters measured at Ulysses are taken into account⁴, a bow shock stand-off distance of $(2.9\text{--}3.7) \times 10^5 \text{ km}$ is expected¹⁴. As solar-wind structures move anti-sunward, they expand, with their dimensions perpendicular to the radial direction increasing linearly with heliocentric distance, which must also be taken into account. Assuming a bow shock flank distance ~ 1.5 times the upstream stand-off distance, a tail diameter of $\sim (9\text{--}11) \times 10^6 \text{ km}$ is expected, which is consistent with the structure at Ulysses.

A lower limit to the tail's actual size can also be estimated from ground-based images. High-quality images of the ion tail were obtained on April 21.8 UT—around two days before the plasma is estimated to have left the comet's head. However, as the comet was already immersed in the fast, uniform solar wind at that time, the tail's diameter probably only varied significantly as a result of changes in the comet's production rate. The visible tail width near the nucleus in the light of H_2O^+ ions was greater than $6.7 \times 10^5 \text{ km}$. This corresponds to a width of greater than $6.3 \times 10^6 \text{ km}$ at Ulysses, which again is similar to that of the structure.

On the basis of plasma instrument¹⁵ measurements, the structure encountered by Ulysses was apparently not pressure-balanced, nor apparently was charge neutrality maintained, as the electron number density remained comparatively stable⁴. As violation of these conditions was unlikely, another source of pressure, such as heavy ions, must have been present; these would have been outside the detection range of the plasma instrument. This suggests that heavy ions of cometary origin were present in the structure. The presence of such ions is confirmed¹⁶ in data returned at the time from the spacecraft's ion composition instrument¹⁷. Decreased broad-band turbulence for ~ 14 hours before the tail crossing is consistent with the tail preventing solar-wind Alfvén wave propagation through it.

Although the crossing's exact location with respect to the tail axis is unknown, some of the tail's parameters can be deduced from its internal magnetic structure. The nested regions were probably due

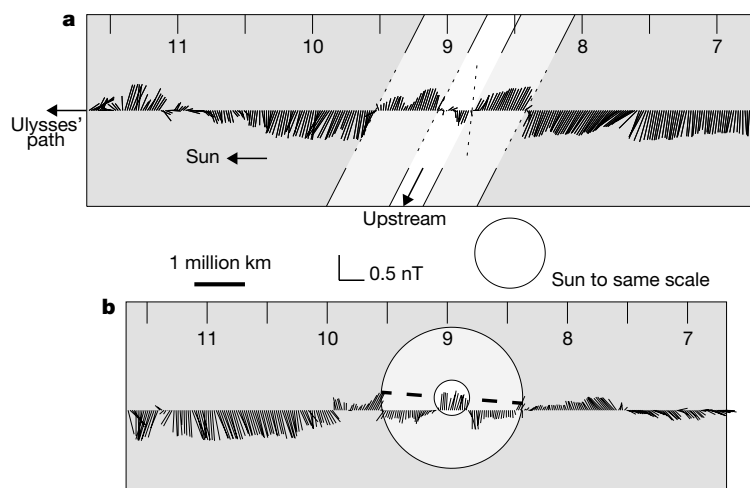


Figure 2 Magnetic field vectors obtained during the tail crossing. Data (averaged over one minute intervals) are presented from 06:40 to 11:40 UT on 1 May 1996, looking perpendicular to the comet's orbital plane (panel **a**), and along the tail's axis (panel **b**). Ulysses entered the tail on its anti-sunward side and left on the sunward side. The Sun's disk to the same scale is shown between the two panels. Vertical ticks mark every half-hour. The scale of magnetic field vectors is also shown. The structure's centre was encountered around 09:00 UT, around which broad symmetry is seen in the vector pattern. The structure is characterized by nested regions of near-constant field orientation, separated by discontinuities (marked by solid lines in **a**, by circles in **b**; a circular cross-section was assumed). In **a**, the discontinuities' orientations, deduced from minimum

variance analysis²⁴, are also shown (dotted lines). Only the clearest of these discontinuities are shown—more are present, possibly extending from $\sim 07:35$ to $10:30$ UT. These concentric structures are interpreted as being different regions of solar-wind plasma, separated by discontinuities distorted during their passage through the comet's sheared flow. Minimum variance analysis²⁴ gives the inferred current sheet orientation shown in **b** (dashed lines); this is consistent with a tail orientated $\sim 63\text{--}70^\circ$ from the solar radial direction. A current sheet was not traversed; however, from the magnetic field vector directions, it is clear that Ulysses passed through the side of the tail nearest the ecliptic. From geometrical considerations, the spacecraft's minimum distance from the tail's axis was greater than $2.5 \times 10^5 \text{ km}$.

to distorted discontinuities propagating through the comet, as were seen at comet Halley¹⁸, but not previously detected in a tail. One of these discontinuities may be related to a solar-wind discontinuity detected by Ulysses at 00:53 UT on 1 May. The observation of such distorted discontinuities is consistent with some tail ray formation theories^{19,20}. The orientations of the solar-wind field lines surrounding the tail suggest upstream pile-up, as the wind encountered the slower tail. Due to the comet's retrograde motion, the tail's orientation was almost opposite to that of the surrounding field (which followed the archimedean spiral expected from the Sun's rotation). This difference in orientations probably meant that a complex interaction region surrounded the tail.

Hyakutake was intrinsically less active than Halley (their water production rates at 0.9 AU from the Sun being 1.8×10^{29} and 5.5×10^{29} molecules s⁻¹, respectively^{11,21}), but Hyakutake's small heliocentric distance on 23 April 1996 resulted in a higher production rate, making it the most productive comet encountered by a spacecraft. This event was unique; only one other unexpected encounter with a comet has been reported²², but that comet remained unidentified. Hyakutake is the first comet to have been encountered when in the fast solar wind.

The presence of a coherent structure so far from the nucleus suggests that the plasma tails of comets may persist as discrete entities for many astronomical units. As the tail of Hyakutake persisted to Ulysses, it is likely that this structure, ultimately produced by an object only ~ 2.4 km across²³, survived to reach the edge of the heliosphere. □

Received 3 November 1999; accepted 8 February 2000.

1. Biermann, L. Kometenschweife und solare Korpuskularstrahlung. *Z. Astrophys.* **29**, 274–286 (1951).
2. Coates, A. J. Ionospheres and magnetospheres of comets. *Adv. Space Res.* **20**, 255–266 (1997).
3. Ip, W.-H. & Axford, W. I. in *Comets* (ed. Wilkening, L. L.) 588–634 (University of Arizona Press, Tucson, 1982).
4. Riley, P., Gosling, J. T., McComas, D. J. & Forsyth, R. J. Ulysses observations of a “density hole” in the high-speed solar wind. *J. Geophys. Res.* **A 103**, 1933–1940 (1998).
5. Balogh, A. *et al.* The magnetic field investigation on the Ulysses mission: Instrumentation and preliminary scientific results. *Astron. Astrophys. Suppl. Ser.* **92**, 221–236 (1992).
6. Brandt, J. C. *et al.* Disconnection events (DEs) in Halley's comet 1985–1986: The correlation with crossings of the heliospheric current sheet (HCS). *Icarus* **137**, 69–83 (1999).
7. Johnstone, A. D. *et al.* Ion flow at comet Halley. *Nature* **321**, 344–347 (1993).
8. Alfvén, H. On the theory of comet tails. *Tellus* **9**, 92–96 (1957).
9. Slavin, J. A. *et al.* Giacobini-Zinner magnetotail: ICE magnetic field observations. *Geophys. Res. Lett.* **13**, 283–286 (1986).
10. Jockers, K. in *Cometary Plasma Processes* (ed. Johnstone, A. D.) 139–152 (Geophysical Monograph 61, American Geophysical Union, Washington DC, 1991).
11. Bertaux, J. L. *et al.* Lyman-alpha observations of comet Hyakutake with SWAN on SOHO. *Planet. Space Sci.* **46**, 555–568 (1998).
12. Jorda, L., Crovisier, J. & Green, D. W. E. in *Asteroids, Comets, Meteors 1991* (eds Harris, A. W. & Bowell, E.) 285–288 (Lunar and Planetary Institute, Houston, 1992).
13. Green, D. W. E. (ed.) *IAU Circ. No. 6388* (1996).
14. Huddleston, D. E., Coates, A. J. & Johnstone, A. D. Predictions of the solar wind interaction with Comet Grigg-Skjellerup. *Geophys. Res. Lett.* **8**, 837–840 (1992).
15. Bame, A. *et al.* The Ulysses solar wind plasma experiment. *Astron. Astrophys. Suppl. Ser.* **92**, 237–265 (1992).
16. Gloeckler, G. *et al.* Interception of comet Hyakutake's ion tail at a distance of 500 million kilometres. *Nature* **404**, 576–578 (2000).
17. Gloeckler, G. *et al.* The solar wind ion composition spectrometer. *Astron. Astrophys. Suppl. Ser.* **92**, 267–289 (1992).
18. Raeder, J., Neubauer, F. M., Ness, N. F. & Burlaga, L. F. Macroscopic perturbations of the IMF by P/Halley as seen by the Giotto magnetometer. *Astron. Astrophys.* **187**, 61–64 (1987).
19. Jones, G. H. *Weak and Strong Comets in the Solar Wind*. Thesis, Univ. London (1997).
20. Rauer, H., Wegmann, R., Schmidt, H. U. & Jockers, K. 3-D MHD simulations of the effect of comoving discontinuities in the solar-wind on cometary plasma tails. *Astron. Astrophys.* **295**, 529–550 (1995).
21. Krankowsky, D. *et al.* In situ gas and ion measurements at comet Halley. *Nature* **321**, 326–329 (1986).
22. Russell, C. T., Luhmann, J. G., Barnes, A., Mihalov, J. D. & Elphic, R. C. An unusual interplanetary event: encounter with a comet? *Nature* **305**, 612–615 (1983).
23. Lisse, C. M. *et al.* The nucleus of comet Hyakutake (C/1996 B2). *Icarus* **140**, 189–204 (1999).
24. Sonnerup, B. U. O. & Cahill, L. J. Magnetopause structure and attitude from Explorer 12 observations. *J. Geophys. Res.* **A 72**, 171–183 (1967).

Acknowledgements

Ulysses research at Imperial College and Queen Mary and Westfield College is supported by the UK Particle Physics and Astronomy Research Council. We thank H. Mikuž for providing images obtained at the Črni Vrh Observatory, Slovenia.

Correspondence and requests for materials should be addressed to G.H.J. (e-mail: g.h.jones@ic.ac.uk).

Interception of comet Hyakutake's ion tail at a distance of 500 million kilometres

G. Gloeckler^{*†}, J. Geiss[‡], N. A. Schwadron[†], L. A. Fisk[†], T. H. Zurbuchen[†], F. M. Ipavich[†], R. von Steiger[‡], H. Balsiger[§] & B. Wilken^{||}

^{*} Department of Physics and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA

[†] Department of Atmospheric, Oceanic, and Space Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA

[‡] International Space Science Institute, Hallerstrasse 6, CH-3012 Bern, Switzerland

[§] Physikalisches Institut, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

^{||} Max-Planck Institut für Aeronomie, D-37189 Katlenburg-Lindau, Germany

Remote sensing observations^{1–5} and the direct sampling of material^{6–8} from a few comets have established the characteristic composition of cometary gas. This gas is ionized by solar ultraviolet radiation and the solar wind to form ‘pick-up’ ions^{9–11}, ions in a low ionization state that retain the same compositional signatures as the original gas. The pick-up ions are carried outward by the solar wind, and they could in principle be detected far from the coma. (Sampling of pick-up ions has also been used to study interplanetary dust^{12,13}, Venus’ tail¹⁴ and the interstellar medium^{15,16}.) Here we report the serendipitous detection of cometary pick-up ions, most probably associated with the tail of

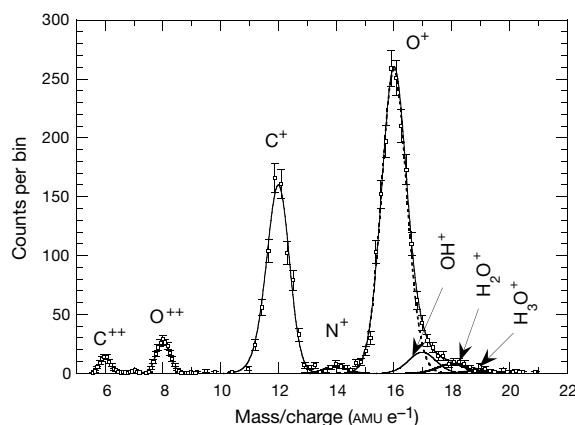


Figure 1 Mass per charge spectrum of C⁺ + O⁺ observed during a one-day period on 1 May 1996. Plotted are the average double and triple coincidence counts of ions per mass/charge (m/q) interval versus the mean m/q of each interval. Only ions with speeds above 0.8 times the simultaneously measured solar wind speed have been included. The coincidence techniques used in SWICS enable us to identify each ion by determining both its mass and mass per charge ratio and measure its speed^{15,16,18}. Resolved ion species include (from left to right) C⁺⁺, N⁺⁺, O⁺⁺, C⁺, N⁺ and O⁺. To obtain the abundance of each ion species we fit gaussian distributions to each of the resolved peaks as well as to the unresolved portion of the spectrum containing molecular ions with masses above 17 AMU. Counts in m/q bins above ~ 17 AMU e⁻¹ are significantly above the dotted curve (the gaussian fit for oxygen) and correspond to singly charged molecular ions with masses of 17, 18 and 19 AMU that are not well separated from each other. The position and width (proportional to the m/q of the ion) of the gaussian distribution corresponding to each ion species is known from best fits to the resolved peaks (C⁺⁺, O⁺⁺ and C⁺). The amplitude of each gaussian was adjusted to minimize χ^2 . The sum of the individual gaussian distributions (solid curve) is an excellent fit to the data. The abundances derived from these best fits (after efficiency corrections) are given in comet Hyakutake in Table 1. Only an upper limit could be obtained for ²⁰Ne⁺.

comet Hyakutake, at a distance of 3.4 AU from the nucleus. Previous observations have provided a wealth of physical and chemical information about a small sample of comets^{6–9}, but this detection suggests that remote sampling of comet compositions, and the discovery of otherwise invisible comets, may be possible.

Cometary ions have been observed during close flybys of comets, at typical distances of less than 10⁷ km (refs 9–11 and 17). It was thought to be unlikely to observe such ions at much larger distances. However, on 1 May 1996 the Solar Wind Ion Composition Spectrometer (SWICS)¹⁸ on Ulysses detected large densities of singly and doubly charged heavy ion species, well above the ubiquitous, steady, low-level pickup ion fluxes^{12,13,15,16}.

The composition of the ionized material we detected (Fig. 1 and comet Hyakutake in Table 1) leaves no doubt that its origin is cometary (compare to comet Halley in Table 1). Following an analysis of the magnetic field, it was concluded that field signatures seen by Ulysses on 1 May were produced by comet Hyakutake¹⁹. In fact, of all known comets only Hyakutake, at ~0.35 AU from the Sun, was positioned such that ionized material from its coma, embedded in the radially outflowing solar wind, would arrive 8 days later, on 1 May, at Ulysses at 3.73 AU. We were greatly surprised to find cometary material so far away from its nucleus. However, given the fortuitous alignment we conclude that the cometary ions we detected most probably came from comet Hyakutake, making it the fourth comet after Giacobini–Zinner, Halley and Grigg–Skjellerup that has been analysed with ion mass spectrometers.

The best mass spectrometer results on coma composition are those obtained from Halley by the Giotto spacecraft; we shall therefore compare the ion abundances we observed from Hyakutake with the ion composition measured in the coma of Halley. There is a remarkable similarity between the abundance patterns of all detected atomic species in the coma of Hyakutake and of Halley. The Hyakutake C⁺/O⁺ ratio is closer to the Halley ratio than the solar C/O ratio. The Hyakutake N⁺/O⁺ ratio is very close to the N⁺/O⁺ ratio of Halley but a factor of 5.5 lower than the solar N/O ratio. The Hyakutake Ne⁺/O⁺ ratio is consistent with that observed in Halley, but is very low compared to the solar Ne/O ratio. It is therefore most unlikely that the ions we observe come from an impulsive solar event spilling predominantly singly charged ions into interplanetary space. Indeed such impulsive solar events have never been observed.

The strong evolution of the composition of water group ions with distance from Halley's nucleus (see Molecules in Table 1) is well understood and agrees with model calculations. We see this pattern continuing for Hyakutake (Table 1) at more than about 10⁶ km from its nucleus. Photo-dissociation of molecular H₂O and CO and the interplay between ion–molecule reactions and dissociative recombination causes fragmentation of molecules and molecular ions until the plasma contains predominantly atomic ions that have a

much longer recombination lifetime than molecular ions. This evolutionary trend, both with distance from the comet and proximity to the Sun, towards a gas of mainly atomic ions is evident in the Hyakutake molecular composition data.

The differences between the Hyakutake and Halley abundances can be easily related to the factor of ~2.6 in the heliocentric distance of these two comets. Dissociation and ionization by EUV in the

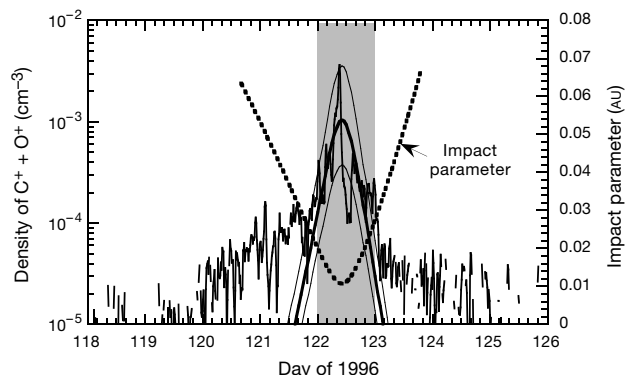


Figure 2 Density of C⁺ + O⁺ ions measured with SWICS at 3.73 AU. At the time of these observations (27 April to 5 May 1996), Ulysses was at 38° latitude in the high-speed (~750 km s⁻¹) solar wind. The dotted curve is a plot of the impact parameter *b*, defined to be the closest distance of comet Hyakutake to the line connecting the Sun and Ulysses. The 750 km s⁻¹ solar wind, and the cometary ions embedded in it, take 7.92 days to travel the distance between the comet and Ulysses; thus the position of Hyakutake 7.92 days earlier is used to compute *b*. The bold curve is the predicted ion density at Ulysses based on a simple model of cometary ion production. Near the comet the density *n*₀ of injected ions may be obtained by balancing the outward convection of cometary ions with the production rate for these ions²⁸: $V_{sw} \times (dn_0/dz) = [Q/(4\pi\lambda_0 r^2)] \times \exp(-r/\lambda_0)$. Here *z* is the distance along the solar wind flow line reaching Ulysses, $r^2 = z^2 + b^2$, *Q* is the gas production rate, and λ_0 is the average ionization mean free path of cometary neutrals. The density of ions near the comet, *n*₀, is computed by integrating *dn*/*dz* from $-\infty$ to the radial position, *R*, of Ulysses. The predicted ion density, *n*, at Ulysses is then reduced by the radial expansion of the solar wind: $n = (r_0/R)^2 n_0$, where *r*₀ is the radial position of Hyakutake. We obtain the best fit to the cometary ion density observed on day 122 (shaded region) using $Q = 2 \times 10^{30}$ s⁻¹, an assumed solar wind speed near the comet (reduced by the mass loading⁹), $V_{sw} = 380$ km s⁻¹ and $\lambda_0 = 1.5 \times 10^6$ km (~0.01 AU). Using a combined production rate inferred for H₂O (ref. 4) and for CO (ref. 3), remote observations yield an estimate for the C⁺ + O⁺ production rate at Hyakutake's radial position (0.35 AU) of ~2.4 × 10³⁰ s⁻¹ (M. Combi, personal communication). The upper and lower thin curves bracketing the observed density profile around the peak are calculated in the same manner as the bold curve except that in computing these curves the value of the impact parameter *b* is changed by ±0.7 × 10⁶ km respectively, corresponding to a deviation from radial solar wind flow by ±0.08°.

Table 1 Abundances of atomic and water group ions in comets

Ratio	Comet Hyakutake* (10 ³ km)	Comet Halley			Solar abundances‡	
		Elements†	Molecules			
			240,000 km	20,000 km¶		3,000 km¶
C ⁺ /O ⁺	0.70 ± 0.06	0.6			0.49	
N ⁺ /O ⁺	0.024 ± 0.005#	0.025#			0.12	
Ne ⁺ /O ⁺	<0.005	<0.008			0.18	
OH ⁺ /O ⁺	0.072 ± 0.015		0.6	1	5	
H ₂ O ⁺ /O ⁺	0.05 ± 0.01		0.2	4	40	
H ₃ O ⁺ /O ⁺	0.015 ± 0.007		<0.02	8	200	
O ⁺ /O ⁺⁺	10 ± 2					
C ⁺ /C ⁺⁺	16 ± 3					

* Present work. For comet Hyakutake so close to the Sun, the atomic ions should correspond to the total coma abundances of C, N and O, because organic compounds, which contain significant fractions of C and N (compare ref. 22), will evaporate much faster at 0.35 AU than at 0.9 AU where comet Halley observations (columns 3–6) were made. Listed uncertainties are a combination of statistical and systematic errors.

† From ref. 23 and 24. Abundances were derived from mass-spectrometric measurements of gas and dust in Halley's coma.

‡ From ref. 27.

|| From ref. 25.

¶ From ref. 26.

A small (compared to solar) N/O ratio is expected. The low nitrogen abundance in Halley, and probably other comets, is thought to be due to loss of N₂ or depletion of N at the time of accretion²³.

outer coma and in the tail of Hyakutake (0.35 AU) should have been basically ~ 7 times faster than at Halley (0.9 AU). This should result in a faster evolution towards a plasma consisting primarily of atomic ions, thus explaining the high abundance of O^+ compared to water group molecular ions in the Hyakutake plasma. H_3O^+ is produced by reactions between ions and H_2O . Its abundance falls off sharply beyond a distance from the nucleus of the comet where the H_2O density has decreased to such low values that ion–molecule reactions become ineffective. Thus, the low value of H_3O^+ observed in Hyakutake is consistent with an origin in the outer part of a coma.

As can be seen from Fig. 2, the maximum $\text{C}^+ + \text{O}^+$ density of 0.0036 cm^{-3} occurred at 9:15 on 1 May (day 122) when the Sun, comet Hyakutake and Ulysses were in closest radial alignment. At that time the shortest distance of Hyakutake to a line connecting the Sun and Ulysses 7.92 days later (the impact parameter shown as the dotted curve; see Fig. 2 legend) reached its minimum of $0.010 \pm 0.006 \text{ AU}$ or $(1.5 \pm 0.9) \times 10^6 \text{ km}$. The uncertainty results from an assumed likely deviation of $\pm 0.1^\circ$ from a strictly radial solar wind flow direction. From a fit (bold curve) to the observed density using a simple model (see Fig. 2 legend) we obtain a gas production rate of $(2 \pm 1) \times 10^{30} \text{ s}^{-1}$ and an ionization mean free path of 0.01 AU for comet Hyakutake at 0.35 AU.

In the coma and tail region the mass density of cometary ions is sufficiently high to significantly deform the magnetic field configuration and decrease the solar wind speed²⁰. A rough calculation using the density near the comet shows that the tail region would intersect the Sun–Ulysses line for impact parameters less than $\sim 0.7 \times 10^6 \text{ km}$, making it quite plausible from this geometrical alignment that Ulysses detected ions from the tail region. The tail crossing is further supported by the observations shown in Fig. 3. At the time of peak density of cometary $\text{C}^+ + \text{O}^+$ ions the solar wind ion densities

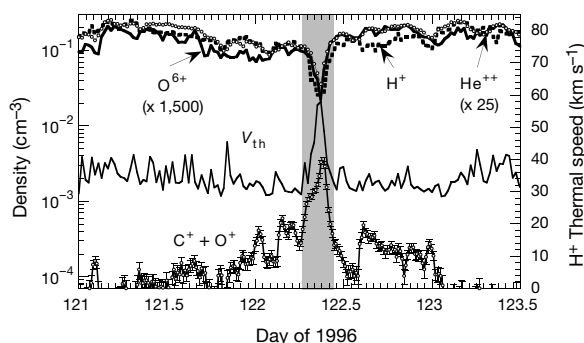


Figure 3 SWICS observations during 60 hours centred on 1 May 1996. Shown are the solar wind H^+ (squares), He^+ ($\times 25$; circles) and O^{6+} ($\times 1,500$; grey line) densities, proton thermal speed (V_{th} ; thin line) and the $\text{C}^+ + \text{O}^+$ ion density (circles with error bars on line). The shaded region indicates the likely traversal of the distant plasma tail of Hyakutake by Ulysses, where the solar wind is significantly deformed due to the pressure of cometary ions. In this region the cometary ion density is larger than the density of solar wind minor ions (He^+ and O^{6+}). Some of the solar wind signatures shown here were previously observed²⁹, but their comet-tail origin was not established until now. A gradual increase in the cometary ion density is evident days before the peak density was reached on day 122.4 and for the next few days the density gradually declined to background level (also see Fig. 2). Rapid variations in the observed $\text{C}^+ + \text{O}^+$ density are notable. This is not surprising as the source of ions at Ulysses appears point-like, and the solar wind velocity is likely to deviate from purely radial flow. The cometary ion density outside the core region is significantly above the maximum predicted density (upper thin curve in Fig. 2) using our simple model. The ions we observe outside the core region were created at larger distances from the comet in the nearly unperturbed solar wind. It is possible that these ions have subsequently undergone substantial diffusion along the magnetic field which smears the point-like cometary ion source over a wide range of impact parameters due to meandering of intertwining magnetic field lines³⁰. Alternatively, more sophisticated modelling of the evolution of cometary molecules in the extended coma may account for the observed excess.

drop abruptly by a factor of ~ 5 and the proton thermal speed doubles. We interpret these solar wind signatures to be additional evidence that Ulysses traversed a portion of the plasma tail of Hyakutake, making this by far the longest comet tail ever detected.

Using a dedicated instrument like SWICS that is capable of detecting atomic and molecular ions up to the mass per charge of CO^+ and S^+ , retains its low background, but has a much larger sensitivity than SWICS, it will now be possible to make a comprehensive search for smaller comets. In this way, the statistics of abundances and orbits of comets can be greatly improved.

The first evidence for an invisible continuously flowing solar wind came from observations of comet tails²¹. The roles are now reversed: pick-up ion composition measurements in the solar wind are a way to detect invisible comets. □

Received 29 December 1999; accepted 11 February 2000.

- Huebner, W. F. & Benkhoff, J. From coma abundances to nucleus composition. *Space Sci. Rev.* **90**, 117–130 (1999).
- Crovisier, J. & Bockelée-Morvan, D. Remote observations of the composition of cometary volatiles. *Space Sci. Rev.* **90**, 19–32 (1999).
- Biver, N. *et al.* Spectroscopic monitoring of comet C/1996 B2 (Hyakutake) with the JCMT and IRAM radio telescopes. *Astrophys. J.* **118**, 1850–1872 (1999).
- Combi, M. R. *et al.* Hubble Space Telescope ultraviolet imaging and high-resolution spectroscopy of water photodissociation products in comet Hyakutake (C/1996 B2). *Astrophys. J.* **494**, 816–821 (1999).
- Lisse, C. M. *et al.* The nucleus of comet Hyakutake (C/1996 B2). *Icarus* **140**, 189–204 (1999).
- von Rosening, T. T., Brandt, J. C. & Farquhar, R. W. The Cometary Explorer mission to comet Giacobini-Zinner. *Science* **232**, 353–356 (1986).
- Encounters with comet Halley: the first results. In *Nature* **321** (Suppl.), 259–366 (1986).
- Altwegg, K., Balsiger, H. & Geiss, J. Composition of the volatile material in Halley's coma from in situ measurements. *Space Sci. Rev.* **90**, 3–18 (1999).
- Ipavich, F. M. *et al.* Comet Giacobini-Zinner - In situ observations of energetic heavy ions. *Science* **232**, 366–369 (1986).
- Gloeckler, G. *et al.* Cometary pick-up ions observed near Giacobini-Zinner. *Geophys. Res. Lett.* **13**, 251–254 (1986).
- Neugebauer, M. *et al.* The density of cometary protons upstream of comet Halley's bow shock. *J. Geophys. Res.* **94**, 1261–1269 (1989).
- Geiss, J., Gloeckler, G., Fisk, L. A. & von Steiger, R. C^+ pickup ions in the heliosphere and their origin. *J. Geophys. Res.* **100**, 23373–23377 (1995).
- Gloeckler, G., Fisk, L. A., Geiss, J., Schwadron, N. A. & Zurbuchen, T. H. Elemental composition of the inner source pickup ions. *J. Geophys. Res.* **105**, 7459–7463 (2000).
- Grünwaldt, H. *et al.* Venus tail ray observation near Earth. *Geophys. Res. Lett.* **24**, 1163–1166 (1997).
- Gloeckler, G. *et al.* Detection of interstellar pickup hydrogen in the solar system. *Science* **261**, 70–73 (1993).
- Gloeckler, G. & Geiss, J. Interstellar and inner source pickup ions observed with SWICS on Ulysses. *Space Sci. Rev.* **86**, 127–159 (1998).
- Balsiger, H. *et al.* Ion composition and dynamics at comet Halley. *Nature* **321**, 330–336 (1986).
- Gloeckler, G. *et al.* The solar wind ion composition spectrometer. *Astron. Astrophys.* **92** (Suppl. Ser.) 267–289 (1992).
- Jones, G. H., Balogh, A. & Horbury, T. S. Identification of comet Hyakutake's extremely long ion tail from magnetic field signatures. *Nature* **404**, 574–576 (2000).
- Ip, W. H. & Axford, W. I. The formation of a magnetic-field-free cavity at Comet Halley. *Nature* **325**, 418–419 (1987).
- Biermann, L. Kometenschweife und solare Korpusskularstrahlung. *Z. Naturforsch.* **7a**, 127–136 (1952).
- Kissel, J. *et al.* Composition of comet Halley dust particles from Giotto observations. *Nature* **321**, 336–337 (1986).
- Geiss, J. Composition measurements and the history of cometary matter. *Astron. Astrophys.* **187**, 859–866 (1987).
- Geiss, J., Altwegg, K., Balsiger, H. & Graf, S. Rare atoms, molecules and radicals in the coma of P/Halley. *Space Sci. Rev.* **90**, 253–269 (1999).
- Neugebauer, M. *et al.* Densities and abundances of hot cometary ions in the coma of P/Halley. *Astrophys. J.* **372**, 291–300 (1991).
- Altwegg, K. *et al.* The ion population between 1300 km and 230000 km in the coma of comet P/Halley. *Astron. Astrophys.* **279**, 260–266 (1993).
- Grevesse, N. & Sauval, A. J. Standard solar composition. *Space Sci. Rev.* **85**, 161–174 (1998).
- Schmidt, H. U. & Wegmann, R. In *Comets* (ed. Wilkening, L. L.) 538–560 (Univ. Arizona Press, Tucson, 1982).
- Riley, P., Gosling, J. T. & McComas, D. J. Ulysses observations of a “density hole” in the high-speed solar wind. *J. Geophys. Res.* **103**, 1933–1940 (1998).
- Jokipii, J. R. & Parker, E. N. Random walk of magnetic lines of force in astrophysics. *Phys. Rev. Lett.* **21**, 44–47 (1968).

Acknowledgements

The SWICS instrument was developed by a collaboration of the universities of Maryland, Bern and Braunschweig, and the Max-Planck Institut für Aeronomie. We thank C. Gloeckler for help with data reduction, and M. R. Combi and M. F. A'Hearn for discussions. This work was supported by NASA/JPL and the Swiss National Science Foundation.

Correspondence should be addressed to G. G. (e-mail: gg10@umail.umd.edu).

A scalable quantum computer with ions in an array of microtraps

J. I. Cirac & P. Zoller

Institute for Theoretical Physics, University of Innsbruck, Austria

Quantum computers require the storage of quantum information in a set of two-level systems (called qubits), the processing of this information using quantum gates and a means of final readout¹. So far, only a few systems have been identified as potentially viable quantum computer models—accurate quantum control of the coherent evolution is required in order to realize gate operations, while at the same time decoherence must be avoided. Examples include quantum optical systems (such as those utilizing trapped ions^{2–9} or neutral atoms^{10–12}, cavity quantum electrodynamics^{13–15} and nuclear magnetic resonance^{16,17}) and solid state systems (using nuclear spins¹⁸, quantum dots¹⁹ and Josephson junctions²⁰). The most advanced candidates are the quantum optical and nuclear magnetic resonance systems, and we expect that they will allow quantum computing with about ten qubits within the next few years. This is still far from the numbers required for useful applications: for example, the factorization of a 200-digit number requires about 3,500 qubits²¹, rising to 100,000 if error correction²² is implemented. Scalability of proposed quantum computer architectures to many qubits is thus of central importance. Here we propose a model for an ion trap quantum computer that combines scalability (a feature usually associated with solid state proposals) with the advantages of quantum optical systems (in particular, quantum control and long decoherence times).

Our model considers ions stored in an array of microtraps²³ which can be realized by electric and/or laser fields (see Fig. 1). Long-lived internal states of the ions carry the qubits. The main idea is to apply an external field to a particular ion such that its motional wave packet is displaced by a small amount depending on the internal state, analogous to the splitting of the atomic wave packet in atom interferometry. By addressing two neighbouring ions, a differential energy shift due to the Coulomb interaction can be induced, conditional on the state of the two qubits. This implements the two-qubit phase gate required for quantum computation. We emphasize that this mechanism differs fundamentally from our proposal for a quantum computer with trapped ions in a linear trap². In that case, the two-qubit gate was based on the exchange of phonons corresponding to the collective centre-of-mass motion of the ions, which initially had to be cooled to zero temperature. In contrast, in our present proposal the motion of the ions is manipulated independently and there is no zero-temperature requirement. This makes the scheme conceptually simpler and obviously scalable.

We first explain qualitatively the operation of our model system, followed by a quantitative analysis of the requirements. We consider a set of N ions confined in independent harmonic potential wells which are separated by some constant distance d (Fig. 1). For simplicity we assume for the moment that each ion is in the ground state of the corresponding potential. We will assume that d is large enough so that: (1) the Coulomb repulsion is not able to excite the vibrational state of the ions; and (2) the ions can be individually addressed. Each ion stores a qubit in two internal, long-lived states, $|0\rangle$ and $|1\rangle$. The initialization and readout of the quantum register can easily be performed as described². Initially, the global wavefunction describing the state of the ions can be factorized into the one corresponding to their internal state and the one corresponding to their motional state. Single-qubit operations on a given ion can be performed by directing a laser beam with the appropriate intensity and phase to the corresponding ion in such a

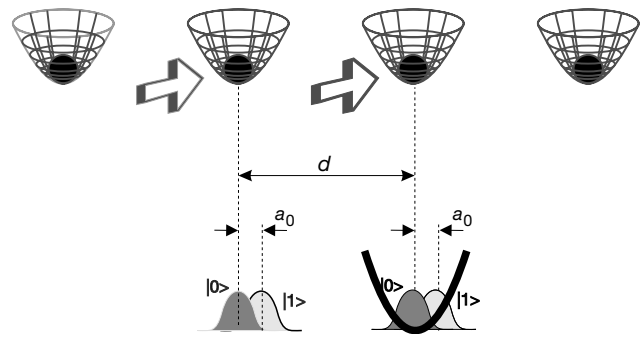


Figure 1 Schematic setup. Ions are trapped in independent traps. For the gate they are pushed provided they are in state $|1\rangle$.

way that its motional state is not affected. Two-qubit gates between two neighbouring ions can be performed by slightly displacing them for a short time T if they are in a particular internal state, say $|1\rangle$. In that case, and provided the ions come back to their original motional state after being pushed, the Coulomb interaction will provide the internal wavefunction (quantum register) with different phases depending on the internal states of the ions. Choosing the time appropriately, the complete process will give rise to the two-qubit gate

$$\begin{aligned} |0\rangle_i |0\rangle_{i+1} |\Psi_{00}\rangle_{\text{rest}} &\rightarrow |0\rangle_i |0\rangle_{i+1} |\Psi_{00}\rangle_{\text{rest}} \\ |0\rangle_i |1\rangle_{i+1} |\Psi_{01}\rangle_{\text{rest}} &\rightarrow |0\rangle_i |1\rangle_{i+1} |\Psi_{01}\rangle_{\text{rest}} \\ |1\rangle_i |0\rangle_{i+1} |\Psi_{10}\rangle_{\text{rest}} &\rightarrow |1\rangle_i |0\rangle_{i+1} |\Psi_{10}\rangle_{\text{rest}} \\ |1\rangle_i |1\rangle_{i+1} |\Psi_{11}\rangle_{\text{rest}} &\rightarrow e^{i\phi} |1\rangle_i |1\rangle_{i+1} |\Psi_{11}\rangle_{\text{rest}} \end{aligned} \quad (1)$$

and in particular, for $\phi = \pi$, we have a phase gate. In general, there will be phase factors $e^{i\phi_i}$ with $i = 1, \dots, 4$ in each line on the right-hand side of equation (1). There is a global phase and (trivial) single particle phases (such as kinetic phases) which can be undone by single-bit operations leaving us with $\phi = \phi_4 - \phi_2 - \phi_3 + \phi_1$. The Ψ symbols denote the internal state of the rest of the ions, and (where we have not written explicitly) the motional state, because it is factorized out at the beginning and at the end of the gate.

To analyse the operating conditions of this scheme in a more quantitative way, we consider two ions 1 and 2 of mass m confined by two harmonic traps of frequency ω in one dimension (see Fig. 1). The generalization of our analysis to three dimensions and N ions is straightforward. We denote by $\hat{x}_{1,2}$ the position operators of the two ions. Now, let us assume that if the ions are in state $|1\rangle$ in addition to the trap, then the i th ion is pushed by a force $F_i(t) = \frac{1}{2} f_i(t) \hbar \omega / a_0$, where $f_i(t)$ is some smooth function of time with $a_0 = (\hbar / 2m\omega)^{1/2}$ the size of the trap ground state. We will take $f_i(t)$ going from 0 to approximately 1 and then back to 0 as time goes from 0 to T . The effect of the force is to shift the centre of the harmonic potential to the position $\bar{x}_i(t) = f_i(t) a_0$. The corresponding potential is

$$V = \sum_{i=1,2} \frac{1}{2} m \omega^2 (\hat{x}_i - \bar{x}_i(t) |1\rangle_i \langle 1|)^2 + \frac{e^2}{4\pi\epsilon_0} \frac{1}{|d + \hat{x}_2 - \hat{x}_1|} \quad (2)$$

The minimum of the potential determines the equilibrium position of the two ions $x_{1,2}^{(0)}$ in the absence of the force $F_i(t)$; we can expand the potential around these equilibrium values, where we conveniently redefine the position of the ions as the displacement around these equilibrium values: $\hat{x}_{1,2} \rightarrow x_{1,2}^{(0)} + \hat{x}_{1,2}$ and $d \rightarrow d + x_2^{(0)} - x_1^{(0)}$. We are interested in the limit where (1) $|\hat{x}_{1,2}| \ll d$ and (2) $\epsilon |\hat{x}_1 \hat{x}_2| / a_0^2 \ll 1$, with ϵ the ratio of the Coulomb energy, $e^2 / (4\pi\epsilon_0 d)$, and the energy of the second ion with respect to the first trap, $\frac{1}{2} m \omega^2 d^2$. Furthermore, we assume that the motional state of the pushed ions will change adiabatically with the potential, which, away from $\epsilon^2 \ll 1$, requires

$|\dot{f}_i(t)| \ll \omega$. Expansion of the Coulomb term in powers of $\hat{x}_{1,2}/d$ gives rise to a term, $-m\omega^2 \hat{x}_1 \hat{x}_2$, in equation (2). It is this term which is responsible for entangling the atoms according to equation (1), giving rise to a conditional phase shift

$$\phi = -\frac{1}{2}\epsilon\omega \int_0^T dt f_1(t)f_2(t) \quad (3)$$

if and only if the internal state of both ions is $|1\rangle$. Equation (3) can simply be interpreted as arising from the energy shifts due to the Coulomb interactions of atoms accumulated on different trajectories according to their internal states

$$\phi = -\frac{e^2}{4\pi\epsilon_0} \int_0^T dt \left[\frac{1}{d + \bar{x}_2 - \bar{x}_1} - \frac{1}{d + \bar{x}_2} - \frac{1}{d - \bar{x}_1} + \frac{1}{d} \right],$$

where the four terms are due to atoms in $|1\rangle_1|1\rangle_2$, $|1\rangle_1|0\rangle_2$, $|0\rangle_1|1\rangle_2$ and $|0\rangle_1|0\rangle_2$, respectively. In summary, by selecting the time of the gate $T \approx 2\pi/(\epsilon\omega)$ (that is, $\phi = \pi$) we have the phase gate, equation (1). We emphasize that equation (3) depends only on mean displacement of the atomic wavepacket and is thus insensitive to the temperature (the width of the wave packet), which will appear in the problem only in higher orders of $x_{1,2}/d$ of our expansion of equation (2), or in cases of non-adiabaticity. Finally, although we have considered here an adiabatic displacement of the trap centre, similar schemes are possible which are based on time-dependent potentials. As an example, in Table 1 we give the values of the relevant parameters of the problem for the case of Ca, at several trapping frequencies.

We now discuss briefly some of the more technical aspects for some physical realizations of the trapping and state-dependent pushing required in our scheme. Regarding trapping, recent developments in nanofabrication techniques allow us to create arrays of electric or magnetic microtraps²³. Alternatively, we can first confine the particles with standard ion traps and then switch on an (additional) off-resonant optical lattice potential so that the ions are confined at the bottom of nonconsecutive wells. For example, with CO₂ lasers, very long confinement times can be obtained (of the order of minutes) without having spontaneous emission, and trap frequencies of tens of MHz can be achieved²⁴. On the other hand, for the pushing we need a force that acts differently depending on the internal state of the ions. Let us consider one simple possibility, in which $|1\rangle$ (or $|0\rangle$) corresponds to one of the $S_{1/2}$ (or $D_{5/2}$) levels, as in Ca, for example (another possibility is to use the method given in ref. 10). A laser standing wave connecting the states $S_{1/2}$ with the $P_{1/2}$, with a detuning Δ and forming some angle θ with respect to the vector that connects the ions participating in the gate, can be used. We denote by $\Omega(t)^2 = f(t)\Omega_0^2$ the square of the Rabi frequency and by $\eta = ka_0\cos(\theta)$ the Lamb–Dicke parameter, where k is the laser wavevector. Choosing $\eta\Omega_0^2/(2\Delta) = \omega$ one achieves the required force $F_i(t) = \frac{1}{2}f_i(t)\hbar\omega/a_0$. On the other hand, the spontaneous emission rate is reduced by a factor $r_f = \Omega_0^2/(2\Delta)^2 \ll 1$ (that is, $\Gamma_{\text{eff}} = r_f\Gamma$). We thus obtain the detunings and Rabi frequencies required: $\Delta = r_f\omega/\eta$ and $\Omega_0 = r_f^{1/2}\omega/\eta$. For a 10-MHz trap frequency with $\eta = 0.1$, we can choose a reduction factor of 10^{-4} (or 10^{-6}) by taking $\Delta = 10^3$ GHz (or 10^5 GHz) and $\Omega_0 = 10$ GHz (or 100 GHz).

Table 1 Experimental parameters for Ca*

	$\epsilon = 0.1$			$\epsilon = 0.01$		
$\omega/2\pi$ (MHz)	1	5	20	1	5	20
d (μm)	12	4	1.6	26	9	3.5
a_0/d ($\times 10^{-3}$)	0.9	1.2	1.5	0.4	0.5	0.7
T (μs)	10	2	0.5	100	20	5

For a given trap frequency ω and a (small) parameter ϵ we give the distance d between the ions, size of the trap ground state relative to the distance a_0/d and time of the gate T . We have taken two values of $\epsilon = 0.1$ and 0.01 so that the condition $\epsilon^2 \ll 1$ is well satisfied. The other requirement $a_0/d \ll 1$ is also clearly fulfilled. The distance between the ions and the corresponding gate times are also displayed. We note that independent traps permits high trap frequencies and therefore short operation times, which presents obvious advantages for avoiding decoherence.

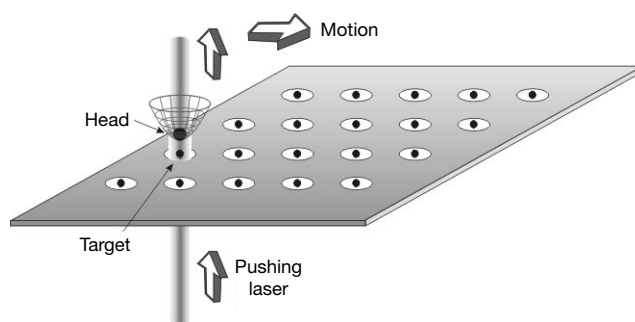


Figure 2 Scalable quantum computer. We envisage a two-dimensional array of independent ion traps²³, and a different ion (Head) that moves above this plane, approaching any particular ion. By switching on a laser propagating in the perpendicular direction to the plane, we can perform the two-qubit gate between the target ion and the head as explained above. In particular, we can swap the state of that ion to the head, which immediately allows us to perform entanglement operations between distant ions.

A possible quantum computer model based on the ideas presented above is outlined in Fig. 2. This scheme has several attractive features: (1) The head and the target can be brought very close together, since we do not have to address them independently. This implies that the gate operation time can be greatly decreased. (2) The distance between the ions in the plane can be very large, since no interaction between them is required. (3) We can think of having several heads running in parallel. (4) We may use a different isotope or ion species for the head, which facilitates the independent motion of the head. (5) The two-dimensional structure facilitates scaling up to a larger number of qubits.

A possible experimental realization of the scalable quantum computer model will represent significant experimental challenges. First, one has to create arrays of microtraps and load them with ions. According to ref. 23, a possible way of doing that is to use elliptical traps, which can be constructed from a single flat electrode, yielding a device that is simpler, smaller and stronger than a linear trap, and is suitable for microfabrication in large arrays. Second, decoherence must be avoided. There are several sources of decoherence in the present model. For example, the motional states of the ions can get entangled with the environment. This is relevant, however, only during the gate operations, and not during the time in which the head is moved (because the internal state of the ions is factorized from the motional degrees of freedom)—provided, of course, that the heating rate is not too high. On the other hand, the time for the gate operation in our scheme (which is only limited by the trapping frequency) can be much smaller than the typical decoherence times reported in experiments with microtraps (ref. 9 and references therein). Finally, laser intensity fluctuations of the pushing laser must be controlled in order to avoid dephasings.

We have here proposed a new scheme to perform quantum computations with trapped ions. Despite being conceptually simple, we believe that the scheme is within reach of present experimental technologies. In particular, our two-dimensional array setup offers an experimental avenue towards reaching a scalable quantum computer. Finally, we would like to stress that apart from using it for quantum computations, our scheme can emulate Ising (and similar) models²⁵ with adjustable parameters simply by leaving on all the pushing lasers which are responsible for effective spin–spin interaction. □

Received 16 November 1999; accepted 16 February 2000.

- DiVincenzo, D. P. Two-bit gates are universal for quantum computation. *Phys. Rev. A* **51**, 1015–1022 (1995).
- Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
- Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. Demonstration of a fundamental quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).

4. Poyatos, J. F., Cirac, J. I. & Zoller, P. Quantum gates with hot trapped ions. *Phys. Rev. Lett.* **81**, 1322–1325 (1998).
5. Steane, A. The ion trap quantum information processor. *Appl. Phys. B* **64**, 623–642 (1997).
6. Turchette, Q. A. *et al.* Deterministic entanglement of two trapped ions. *Phys. Rev. Lett.* **81**, 3631–3634 (1998).
7. Roos, Ch. *et al.* Quantum state engineering on an optical transition and decoherence in a Paul trap. Pre-print quant-ph/9909038 at (xxx.lanl.gov) (1999).
8. Sorensen, A. & Molmer, K. Quantum computation with ions in thermal motion. *Phys. Rev. Lett.* **82**, 1971–1974 (1999).
9. Wineland, D. J. *et al.* Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. Natl Inst. Stand. Technol.* **103**, 259 (1998).
10. Jaksch, D., Briegel, H.-J., Cirac, J. I., Gardiner, C. W. & Zoller, P. Entanglement of atoms via cold controlled collisions. *Phys. Rev. Lett.* **82**, 1975–1978 (1999).
11. Brennen, G. K., Caves, C. M., Jessen, P. S. & Deutsch, I. H. Quantum logic gates in optical lattices. *Phys. Rev. Lett.* **82**, 1060–1063 (1999).
12. Calarco, T. *et al.* Quantum gates with neutral atoms: Controlling collisional interactions in time dependent traps. Pre-print quant-ph/9905013 at (xxx.lanl.gov) (1999).
13. Pellizzari, T., Gardiner, S. A., Cirac, J. I. & Zoller, P. Decoherence, continuous observation, and quantum computing: a cavity QED model. *Phys. Rev. Lett.* **75**, 3788–3791 (1995).
14. Turchette, Q. A., Hood, C. J., Lange, W., Mabuchi, H. & Kimble, H. J. Measurement of conditional phase shifts for quantum logic. *Phys. Rev. Lett.* **75**, 4710–4713 (1995).
15. Maître, X. *et al.* Quantum memory with a single photon in a cavity. *Phys. Rev. Lett.* **79**, 769–772 (1997).
16. Cory, D. G., Fahmy, A. F. & Havel, T. F. Ensemble quantum computing by NMR spectroscopy. *Proc. Natl Acad. Sci. USA* **94**, 1634–1639 (1997).
17. Gershenfeld, N. A. & Chuang, I. L. Bulk spin-resonance quantum computation. *Science* **275**, 350–356 (1997).
18. Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
19. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
20. Makhlin, Y. & Schön, G. Josephson-junction qubits with controlled couplings. *Nature* **398**, 305–307 (1999).
21. Beckman, D., Chari, A. N., Devabhaktuni, S. & Preskill, J. Efficient networks for quantum factoring. *Phys. Rev. A* **54**, 1034–1063 (1996).
22. Steane, A. M. Efficient fault-tolerant quantum computing. Pre-print quant-ph/9809054 at (xxx.lanl.gov) (1998).
23. DeVoe, R. G. Elliptical ion traps and trap arrays for quantum computation. *Phys. Rev. A* **58**, 910–914 (1998).
24. Friebe, S., D'Andrea, C., Walz, J., Weitz, M. & Hänsch, T. W. CO₂ laser optical lattice with cold rubidium atoms. *Phys. Rev. A* **57**, R20–R23 (1998).
25. Lloyd, S. Universal quantum simulators. *Science* **273**, 1073–1078 (1996).

Acknowledgements

We thank R. Blatt, D. Leibfried and D. Wineland for comments. This work was supported by the Austria Science Foundation, TMR networks from the European Community, and the Institute for Quantum Information GmbH.

Correspondence and requests for materials should be addressed to P.Z.
(e-mail: peter.zoller@uibk.ac.at).

Magnetoresistance from quantum interference effects in ferromagnets

N. Manyala*, Y. Sidis*†, J. F. DiTusa*, G. Aeppli‡, D.P. Young§ & Z. Fisk§

* Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803, USA

† NEC, 4 Independence Way, Princeton, New Jersey 08540, USA

§ National High Magnetic Field Facility, Florida State University, Tallahassee, Florida 32306, USA

The desire to maximize the sensitivity of read/write heads (and thus the information density) of magnetic storage devices has stimulated interest in the discovery and design of new magnetic materials exhibiting magnetoresistance. Recent discoveries include the 'colossal' magnetoresistance in the manganites^{1–4} and the enhanced magnetoresistance in low-carrier-density ferromagnets^{4–6}. An important feature of these systems is that the electrons involved in electrical conduction are different from those responsible for the magnetism. The latter are localized

and act as scattering sites for the mobile electrons, and it is the field tuning of the scattering strength that ultimately gives rise to the observed magnetoresistance. Here we argue that magnetoresistance can arise by a different mechanism in certain ferromagnets—quantum interference effects rather than simple scattering. The ferromagnets in question are disordered, low-carrier-density magnets where the same electrons are responsible for both the magnetic properties and electrical conduction. The resulting magnetoresistance is positive (that is, the resistance increases in response to an applied magnetic field) and only weakly temperature-dependent below the Curie point.

The oxides of manganese that are known to have high magnetoresistance are derived from chemical doping of insulators that are also magnetically ordered. Thus, the local moments that are ordered in the doped materials already exist in the insulator, with the result that, to a first approximation, metallicity and magnetism are independent properties⁶. The electrical resistance is reduced when there is less scattering of the conduction electrons by the spins of the localized electrons. To the extent that an external field reduces the disorder among the local spins, the scattering will be reduced, resulting in a negative magnetoresistance. When searching for different mechanisms for magnetoresistance, we need to consider compounds where the insulating parent is non-magnetic. In addition, the parent should have strong electron–electron interactions so that magnetism appears readily upon doping. Insulators which satisfy this criterion are referred to as strongly correlated, or Kondo insulators⁷. A particularly simple Kondo insulator is FeSi, which has attracted attention for over 30 years because of its anomalous temperature-dependent electrical, optical and magnetic properties^{7–11}. Previous investigations reveal a metal–insulator transition with either Al substitution for Si (hole doping) or Co substitution for Fe (electron doping) at the level of 1% (refs 12 and 13). The transition is very similar to that found in classic semiconductors such as phosphorus-doped silicon or antimony-doped germanium, except that the effective masses of the holes (electrons) are much larger. FeSi is isostructural (cubic B-20) and continuously miscible with the classic metallic ferromagnet MnSi (see, for example, ref. 14) and the diamagnetic metal CoSi (ref. 8); it is therefore possible to control the carrier sign—positive (holes) for substitution of Mn for Fe and negative (electrons) for substitution of Co—and the density throughout a magnetically interesting phase diagram, shown in Fig. 1. Modest (20%) dilution of Mn by Fe destroys the ferromagnetism of MnSi while retaining metallicity. In contrast, the alloy Fe_{1–y}Co_ySi is remarkable in that although the two end members ($y = 0$) and ($y = 1$) are both non-magnetic, it is magnetic for almost all intermediate compositions^{8,15,16,17}. This is the alloy that displays unusual magnetoresistance, clearly distinguishable from that found in more ordinary magnets such as MnSi.

Our samples were either polycrystalline pellets or small single crystals grown from Sb and Sn fluxes. We produced polycrystalline samples from high purity starting materials by arc melting in an argon atmosphere. To improve homogeneity, the Fe_{1–y}Co_ySi (Fe_{1–x}Mn_xSi) samples were annealed for 24 hours at 1,200 °C (four days at 1,000 °C) in evacuated quartz ampoules. Powder X-ray spectra showed all samples to be single phase with a lattice constant that is linearly dependent on Co and Mn concentration. The linearity demonstrates that Co or Mn successfully replaces Fe over the entire concentration range ($0 \leq x \leq 1$; $0 \leq y \leq 1$). Energy dispersive X-ray microanalysis yielded results consistent with the nominal concentrations. The Hall effect data establish the sign and density of the carriers in our samples and demonstrate a systematic increase in carrier density (n) proportional to the Co and Mn concentration at small x and y . They clearly show the existence of a metallic state in the concentration range studied ($0.03 \leq x \leq 1$; $0.05 \leq y \leq 0.30$), with only a small variation of n with temperature (T). In particular, n does not appear to change as the Curie temperature (T_C) is traversed.

† Present address: Laboratoire Léon Brillouin, CEA-CNRS, CE Saclay, 91191 Gif-Sur-Yvette, France.

Figure 1 summarizes the low-field measurements of the resistivity (ρ) and magnetization (M) throughout the alloy series. Figure 1a shows that the generic behaviour is metallic (increasing conductivity with decreasing T) except in a narrow sliver of temperature–composition space around pure, insulating FeSi. Figure 1b, displaying M measured at 1 T, shows how spin order, with Curie temperatures of 50 K and below, is found in two distinct regions of the phase diagram. It is surprising that substitution of Mn—with its great tendency towards magnetism, as demonstrated by the order found in the end member, MnSi—yields behaviour not very different from substitution of Al for Si (ref. 12). Thus, at low doping, Mn is like Al, in that it inserts holes into the valence band of FeSi. That the outcome is largely independent of the site used to introduce holes represents a great and unexpected simplification of the physics of FeSi and its derivatives. Figure 1c gives the saturation magnetizations (P_S) in the ordered states. The result is that the polarization rises in nearly linear fashion with low Co concentration, with a constant of proportionality which is $1 \pm 0.12 \mu_B$ per Co ion¹⁸. Because the Hall effect yields one electron per Co ion in the same regime, it seems that we have a fully spin-polarized electron gas. In contrast, in the weak itinerant ferromagnet MnSi, we see a substantially reduced moment of $0.43 \pm 0.05 \mu_B$ per Mn ion.

We now discuss our main result: a new type of magnetoresistance in the metallic magnets found on the Co-doped side of the phase diagram. Figure 2a and c show temperature-dependent resistivities measured in various external fields for $\text{Fe}_{0.8}\text{Co}_{0.2}\text{Si}$, where $T_C = 35$ K,

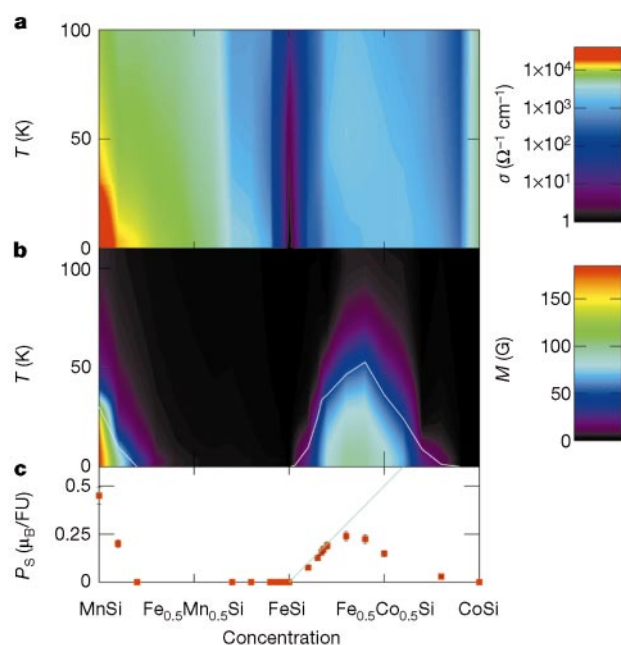


Figure 1 Phase diagram of $\text{Fe}_{1-x}\text{Mn}_x\text{Si}$ and $\text{Fe}_{1-x}\text{Co}_x\text{Si}$. **a**, Conductivity σ versus temperature and concentration. **b**, Magnetization M at 1 T versus temperature and concentration. White line represents the Curie temperature in zero field. **c**, Spontaneous magnetization (P_S) as determined from the saturated value of the magnetization at high magnetic field versus concentration. FU, formula unit. Light-blue line represents behaviour at $1 \mu_B$ per Co ion. The electrical conductivity was measured on rectangular samples with thin platinum wires attached with silver paste. We collected transverse and longitudinal magnetoresistance and Hall effect data at 19 Hz using lock-in techniques. Magnetization measurements were made between 1.7 and 400 K in fields from -5.0 to 5.0 T in a SQUID magnetometer and from 0 to 32 T in a vibrating reed magnetometer at the National High Magnetic Field Laboratory. Technically all of the magnetic materials investigated in our experiments are helimagnets: ferromagnets with long-period helical spin structure. However, in these materials the carrier mean free-paths are much shorter than the helix period^{15,28}. Therefore, for the purposes of discussing the carrier transport, they are essentially ferromagnetic and are referred to as such.

as well as for MnSi, with a similar T_C of 30 K, whereas Fig. 2b and d show the corresponding magnetoresistance ratios $\Delta\rho/\rho$. The differences between the compounds are striking. The MnSi data display the temperature dependence expected for an itinerant ferromagnet¹⁴. Near T_C , $d\rho/dT$ has a peak, due to carrier scattering from spin fluctuations^{8,19}. A magnetic field (H) suppresses the scattering, resulting in the magnetoresistance¹⁹ shown in Fig. 2d: $(\rho(T, H) - \rho(T, 0))/\rho(T, 0)$. A strong negative magnetoresistance (over 20%) occurs only around T_C (ref. 20). Our $\text{Fe}_{1-x}\text{Co}_x\text{Si}$ samples display a zero-field ρ an order of magnitude larger than that of the MnSi sample, but it becomes comparable when multiplied by the number of carriers. The difference between MnSi and $\text{Fe}_{1-x}\text{Co}_x\text{Si}$ is that ρ increases continuously with T below T_C (ref. 15). Furthermore, a magnetic field does not suppress the increase, but instead enhances it, not only near T_C but for all $T < T_C$. Thus, the magnetoresistance in $\text{Fe}_{1-x}\text{Co}_x\text{Si}$ is positive; and, although it turns on near T_C as well as showing a small peak there, its existence is not tied to the magnetic critical fluctuations found near the Curie point. The magnetoresistance shown in Fig. 2b can be as large as 7% near T_C and falls off gradually below this temperature. The effect is enhanced as n is decreased, growing with proximity to the metal–insulator transition with our 10% sample having the largest magnetoresistance (10% at 4 K and 5 T).

Figure 3 shows the field dependence of ρ and M for $\text{Fe}_{0.85}\text{Co}_{0.15}\text{Si}$ at a variety of temperatures. As T is reduced below $T_C = 25$ K, the magnetization varies almost discontinuously with H near 0, exactly as expected for this nearly ferromagnetic helimagnet (see Fig. 1 legend). At the lowest T , the saturation value of M is approached almost immediately for $H > 0$ and $H < 0$. The magnetoresistance ratio $\Delta\rho/\rho$ does not behave analogously: it grows with H at a rate weakly dependent on T . So, in contrast to what occurs for essentially all magnetoresistive materials ranging from MnSi to the many oxides of manganese, the magnetization M is very far from being the dominant controlling factor of ρ . Even so, it is clear that M does have a role to play for small H : above T_C where M is a smooth

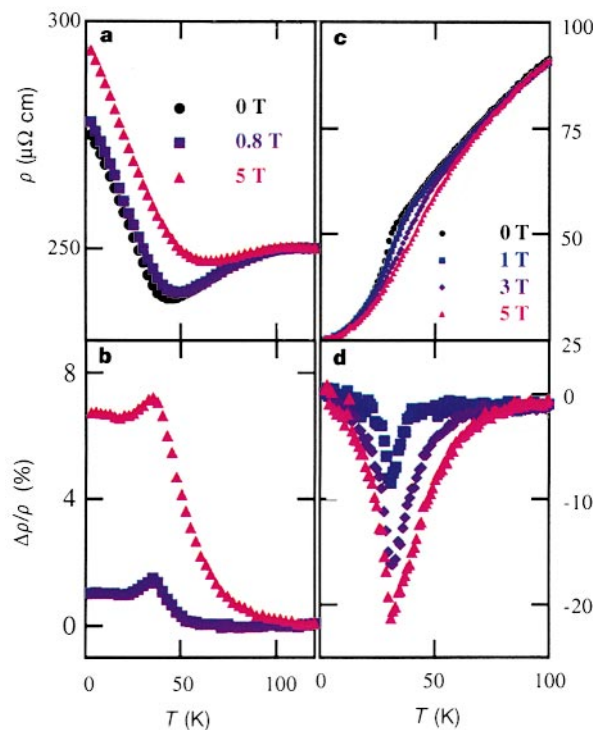


Figure 2 Temperature dependence. **a–d**, The resistivity ρ (**a**, **c**) and magnetoresistance $\Delta\rho/\rho$ (**b**, **d**) for crystalline $\text{Fe}_{0.8}\text{Co}_{0.2}\text{Si}$ (**a**, **b**) and polycrystalline MnSi (**c**, **d**) versus temperature. Magnetic fields are given in the figure.

(linear) function of H , the magnetoresistance is a parabolic function of H , whereas below T_C , when $M(H)$ appears to be singular near $H = 0$, the magnetoresistance acquires a cusp-like minimum near $H = 0$.

Mechanisms for the unusual magnetoresistance in $\text{Fe}_{1-x}\text{Co}_x\text{Si}$ can involve grain boundaries, orbital effects, or electron spins in the bulk of the material. Because we find positive and weakly temperature-dependent magnetoresistance of comparable magnitude in our single crystals, grain boundaries are unlikely to play any role²¹. Furthermore, the similarity of longitudinal and transverse magnetoresistance (Fig. 3a) make orbital effects, which would contribute in the transverse but not the longitudinal geometry, a minor contributor to the magnetoresistance (see, for example, ref. 22; ref. 23).

This leaves coupling to the bulk electron spins as the most likely cause of the anomalous magnetoresistance. Figure 4a and b suggests aspects of the microscopic origin of the observations. First, the conductivity ($\sigma = 1/\rho$) is T - and H -dependent down to the lowest temperatures measured (200 mK). Furthermore, σ is well described by a $T^{1/2}$ dependence, whereas for large H , $\sigma(H, T)$ approaches an

asymptote roughly proportional to $H^{1/2}$.

Has behaviour of the type shown in Fig. 4a and b been seen before? In classic paramagnetic materials, theory has predicted and exhaustive experiments have shown that near the metal–insulator transition the Coulomb interactions between carriers are enhanced by the diffusive nature of the transport^{24–26}. The enhancement results from the finite probability of two carriers interacting more than once within a phase-breaking scattering time²⁴. The scattering amplitudes interfere coherently, leading to an increased Coulomb coupling and a square-root singularity in the electronic density-of-states at the Fermi energy. The resulting $\sigma(H, T)$, which measures the energy-dependent density-of-states either thermally or by the Zeeman effect, has $H^{1/2}$ (for $T = 0$) and $T^{1/2}$ dependences^{24,25}. Our experiments on $\text{FeSi}_{1-x}\text{Al}_x$ have demonstrated that such electron–electron interaction effects dominate the low T transport¹². For comparison we have included in Fig. 4a and b our data for $\text{FeSi}_{0.95}\text{Al}_{0.05}$ and $\text{Fe}_{0.92}\text{Mn}_{0.08}\text{Si}$ which have similar n and σ , but remain paramagnetic at low T . We can see that the T dependence for the two samples is similar. Indeed, for $H = 9$ T, a field chosen to be well above that required to saturate M up to high temperatures in $\text{Fe}_{1-x}\text{Co}_x\text{Si}$, the prefactors of the $T^{1/2}$ term are within 30% of each other for the very different dopants. The field dependences are also similar approaching asymptotes proportional to $H^{1/2}$, although the amplitudes of the asymptotes are more diverse, but still remain within factors of three of each other for the different dopants.

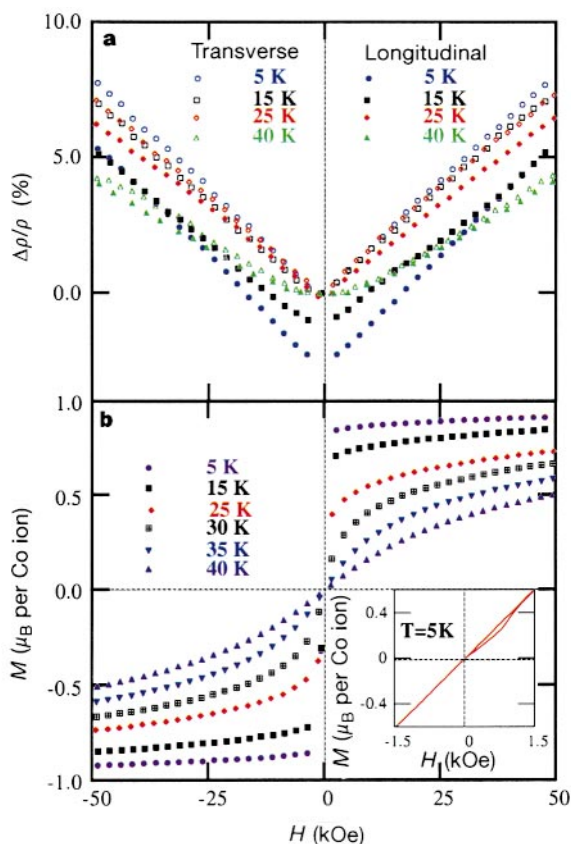


Figure 3 Field dependence of magnetoresistance and magnetization. **a**, Longitudinal and transverse magnetoresistance of $\text{Fe}_{0.85}\text{Co}_{0.15}\text{Si}$ at various temperatures. **b**, Magnetization of $\text{Fe}_{0.85}\text{Co}_{0.15}\text{Si}$ at various temperatures. Inset, history dependence of the magnetization showing a small hysteresis upon increasing the field from zero. The sample was initially cooled to 5.0 K in zero field and the magnetization was measured at positive increasing fields and then cycled through negative fields to zero. At low H and low T the longitudinal magnetoresistance is negative, probably as a result of the anomalous magnetoresistance common to ferromagnets²². This magnetoresistance results from strong spin-orbit coupling in much the same way as the anomalous Hall effect does and is dependent on the orientation between $\mathbf{M}$ and the current density, $\mathbf{J}$. After subtraction of this low-field anisotropic magnetoresistance there are only slight differences between the transverse and longitudinal magnetoresistance. Thus, we conclude that there is little contamination of our data from an ordinary Lorentz magnetoresistance which is anisotropic and dependent on the orientation of $\mathbf{J}$ and $\mathbf{B}$ (ref. 22). The symmetry of our magnetoresistance data about $H = 0$ shows that there is negligible contamination from the (asymmetric) Hall effect.

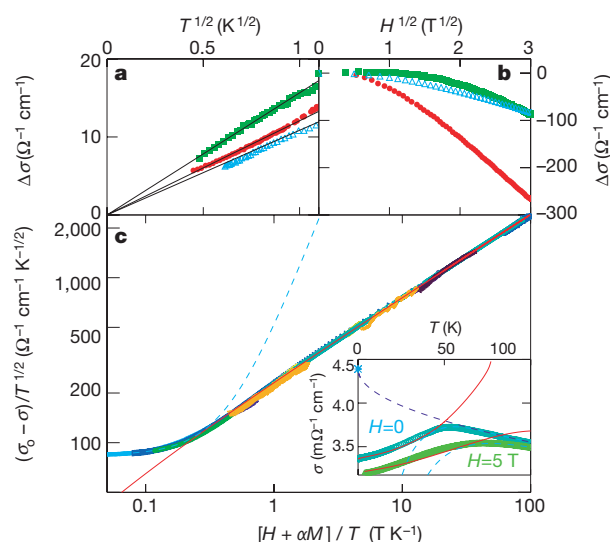


Figure 4 Magnetotransport. **a**, Magnetoconductivity of $\text{Fe}_{0.9}\text{Co}_{0.1}\text{Si}$ (red circle), $\text{FeSi}_{0.95}\text{Al}_{0.05}$ (blue triangle) and $\text{Fe}_{0.92}\text{Mn}_{0.08}\text{Si}$ (green square). Change in conductivity, $\Delta\sigma = \sigma(H, T) - \sigma(H, 0)$ determined from fits of the data to a $T^{1/2}$ dependence, versus $T^{1/2}$ at $H = 9$ T. $\sigma(0, 0) = 1,910 \Omega^{-1} \text{cm}^{-1}$ for $\text{Fe}_{0.9}\text{Co}_{0.1}\text{Si}$, $1,260 \Omega^{-1} \text{cm}^{-1}$ for $\text{FeSi}_{0.95}\text{Al}_{0.05}$ and $1,540 \Omega^{-1} \text{cm}^{-1}$ for $\text{Fe}_{0.92}\text{Mn}_{0.08}\text{Si}$. **b**, $\Delta\sigma = \sigma(H, T) - \sigma(0, T)$ versus $H^{1/2}$ at $T = 0.25$ K. Symbols represent the same samples as in **a**. **c**, Scaling plot of the conductivity $(\sigma - \sigma_0)/T^{1/2}$ versus H_{eff}/T for $\text{Fe}_{0.9}\text{Co}_{0.1}\text{Si}$, with H_{eff} taken as $H + \alpha M$ and with σ_0 and α determined by the best scaling of all our T - and H -dependent data. The data shown include temperature sweeps at constant fields of 0 (0.2–100 K, teal circle), 0.8 T (2–100 K, light-blue filled circle), 3 T (2–100 K, dark-blue forward arrow) and 5 T (2–100 K, green upright cross). Also shown are constant-temperature field sweeps at temperatures of 0.3 K (0–9 T, dark-blue square), 1.2 K (0–32 T, black backward arrow), 1.5 K (0–9 T, purple dotted circle), 4 K (0–32 T, orange diamond), 5 K (0–5 T, yellow-green asterisk), 15 K (0–5 T, yellow cross) and 30 K (0–5 T, violet triangle, and 0–32 T, red inverted triangle). Light-blue dashed line represents a fit to the data for $H_{\text{eff}}/T < 0.25$ in the form: $a + b(H_{\text{eff}}/T)^2$. Solid red line represents a fit to the data for $H_{\text{eff}}/T > 0.25$ in the form: $c + d(H_{\text{eff}}/T)^{1/2}$. Inset, red line and dashed light-blue line represent the same fits as in the main part of the figure. The dashed purple line represents the zero H_{eff} conductivity in our model. Light-blue asterisk represents σ_0 , the zero-temperature, zero-magnetic-field value of the conductivity determined from the scaling of the data.

Given the similarities between the data for paramagnetic Al- and Mn-doped FeSi and ferromagnetic $\text{Fe}_{1-y}\text{Co}_y\text{Si}$, and because the parameters found in the detailed comparison between theory and experiment for $\text{FeSi}_{1-z}\text{Al}_z$ are reasonable, we propose that the low-temperature magnetotransport in $\text{Fe}_{1-y}\text{Co}_y\text{Si}$ is due to electron-electron interactions enhanced by disorder. The proposal is phenomenological, so that its success in describing the data does not exclude underlying physical explanations different from ours.

The standard theory for paramagnetic disordered metals usefully encapsulates $\sigma(H, T)$ by $(\sigma - \sigma_0)/T^{1/2} = f(g\mu_B H/k_B T)$ where $f(x)$ is a scaling function whose limiting form is x^2 for $x \ll 1$ and $x^{1/2}$ for $x \gg 1$ (refs 24 and 25). Because $\text{Fe}_{1-y}\text{Co}_y\text{Si}$ is, to our knowledge, the first ordered ferromagnet for which the $T^{1/2}$ and $H^{1/2}$ terms are present, no theory is available for ferromagnets. Even so, it seems reasonable to believe that the main difference between paramagnets and ferromagnets is simply that for the ferromagnet, in addition to the external field, there is a large spontaneous field due to the ordered moment. Thus, the effective field is really $H_{\text{eff}} = H + \alpha M$ (where α is a constant) rather than H alone. We then imagine that for the ferromagnet, we should simply insert H_{eff} where H appears in the expressions for $\sigma(H, T)$ derived for disordered paramagnets with electron-electron interactions²⁷.

We have checked whether the scaling theory, posited for the paramagnets and generalized by the replacement of H by H_{eff} , works for ferromagnetic $\text{Fe}_{1-y}\text{Co}_y\text{Si}$. Figure 4c shows the outcome, where we have plotted all of our temperature- and field-dependent data below 100 K for the $x = 0.30$ sample in the form $(\sigma - \sigma_0)/T^{1/2}$ versus H_{eff}/T with $\alpha = 1,900$ and $\sigma_0 = 4,400 \Omega^{-1} \text{cm}^{-1}$ chosen to minimize the difference of the data from a piecewise linear form. Note that σ_0 represents the zero- H_{eff} , zero- T conductivity shown in the inset to Fig. 4c. The data scale well and result in a $(H_{\text{eff}}/T)^2$ form for $H_{\text{eff}}/T < 0.25 \text{ T K}^{-1}$ (dashed blue line) and a $(H_{\text{eff}}/T)^{1/2}$ form for $H_{\text{eff}}/T > 0.25 \text{ T K}^{-1}$ (solid red line). The data for the $y = 0.1, 0.15$ and 0.20 samples scale equally well. The inset to Fig. 4c shows the $H = 0$ and 5 T conductivity of the $y = 0.30$ sample, along with the simple asymptotes extracted from the scaling analysis of the main part of Fig. 4, with H_{eff} calculated using the bulk magnetometry data. We conclude from the quality of both the scaling and the fits that, outside of small low-field effects, the field and T -dependent conductivity up to 100 K appears entirely determined by a square-root singularity in the density-of-states probably associated with enhanced electron-electron interactions in a disordered low n ferromagnet. Until now, such $E^{1/2}$ terms have been deemed relevant only at ultra-low temperatures ($< 1 \text{ K}$). Thus, our discovery of a positive magnetoresistance not only shows a new mechanism for magnetoresistance in ferromagnets, but also demonstrates the relevance of quantum interference effects to electrical properties measured well outside the sub-Kelvin domain of low-temperature physics. $\square$

Received 21 October 1999; accepted 1 February 2000.

- Searle, C. W. & Wang, S. T. Studies of the ionic ferromagnet (LaPb)MnO₃. III Ferromagnetic resonance studies. *Can. J. Phys.* **47**, 2703–2708 (1969).
- Jin, S. *et al.* Thousandfold change change in the resistivity in magnetoresistive La-Ca-Mn-O films. *Science* **264**, 413–415 (1994).
- Tokura, Y. *et al.* Origins of colossal magnetoresistance in perovskite-type manganese oxides. *J. Appl. Phys.* **79**, 5288–5291 (1996).
- Shimakawa, Y., Kubo, Y. & Manako, T. Giant magnetoresistance in $\text{Ti}_2\text{Mn}_2\text{O}_7$ with the pyrochlore structure. *Nature* **379**, 53–55 (1996).
- Ramirez, A. P. & Subramanian, M. A. Large enhancement of magnetoresistance in $\text{Ti}_2\text{Mn}_2\text{O}_7$: pyrochlore versus perovskite. *Science* **277**, 546–549 (1997).
- Majumdar, P. & Littlewood, P. B. Dependence of magnetoresistivity on charge-carrier density in metallic ferromagnets and doped magnetic semiconductors. *Nature* **395**, 479–481 (1998).
- Aeppli, G. & Fisk, Z. Kondo insulators. *Comments Condens. Matter Phys.* **16**, 155–165 (1992).
- Wernick, J. H., Wertheim, G. K. & Sherwood, R. C. Magnetic behavior of the monosilicides of the 3d-transition elements. *Mater. Res. Bull.* **7**, 1431–1441 (1972).
- Jaccarino, V., Wertheim, G. K., Wernick, J. H., Walker, L. R. & Arajs, S. Paramagnetic excited state of FeSi. *Phys. Rev.* **160**, 476–482 (1967).
- Schlesinger, Z. *et al.* Unconventional charge gap formation in FeSi. *Phys. Rev. Lett.* **71**, 1748–1751 (1993).
- van der Marel, D., Damaschelli, A., Schulte, K. & Menovsky, A. A. Spin, charge, and bonding in transition metal mono-silicides. *Physica B* **244**, 138–147 (1998).

- DiTusa, J. F. *et al.* Metal-insulator transitions in the Kondo insulator FeSi and classic semiconductors are similar. *Phys. Rev. Lett.* **78**, 2831 (1997); erratum **78**, 4309 (1997).
- Chernikov, M. A. *et al.* Low-temperature transport, optical, magnetic, and thermodynamic properties of $\text{Fe}_{1-x}\text{Co}_x\text{Si}$. *Phys. Rev. B* **56**, 1366–1375 (1997).
- Moriya, T. *Spin Fluctuations in Itinerant Electron Magnetism* (ed. Fulde, P.) (Springer, Berlin/Heidelberg/New York/Tokyo, 1985).
- Beille, J., Voiron, J. & Roth, M. Long period helimagnetism in the cubic B20 $\text{Fe}_3\text{Co}_{1-x}\text{Si}$ and $\text{Co}_2\text{Mn}_{1-x}\text{Si}$ alloys. *Solid State Commun.* **47**, 399–402 (1983).
- Ishimoto, K., Ohashi, M., Yamauchi, H. & Yamaguchi, Y. Itinerant electron ferromagnetism in $\text{Fe}_{0.8}\text{Co}_{0.2}\text{Si}$ studied by polarized neutron diffraction. *J. Phys. Soc. Jpn* **61**, 2503–2511 (1992).
- Kawarazaki, S., Yasuoka, H., Nakamura, Y. & Wernick, J. H. Magnetic properties of $(\text{Fe}_{1-x}\text{Co}_x)\text{Si}$. *J. Phys. Soc. Jpn* **41**, 1171 (1976).
- Paschen, S., Pushin, D., Ott, H. R., Young, D. P. & Fisk, Z. Magnetism and electrical transport in $\text{Fe}_{0.9}\text{TM}_{0.1}\text{Si}$, TM = Co, Rh, Ru. *Physica B* **259–261**, 864–865 (1999).
- Ueda, K. Effect of a magnetic field on spin fluctuations in weakly ferromagnetic metals. *Solid State Commun.* **19**, 965–968 (1976).
- Kadowaki, K., Okuda, K. & Date, M. Magnetization and magnetoresistance of MnSi. *J. Phys. Soc. Jpn* **51**, 2433–2438 (1982).
- Hwang, H., Cheong, S.-W., Ong, N. P. & Batlogg, B. Spin-polarized intergrain tunneling in $\text{La}_{2/3}\text{Sr}_{1/3}\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 2041–2044 (1996).
- Campbell, I. A. & Fert, A. *Ferromagnetic Materials* Vol. 3 (ed. Wohlfarth, E. P.) 751–804 (North-Holland, Amsterdam, 1982).
- Xu, R. *et al.* Large magnetoresistance in non-magnetic silver chalcogenides. *Nature* **390**, 57–60 (1997).
- Lee, P. A. & Ramakrishnan, T. V. Disordered electronic systems. *Rev. Mod. Phys.* **57**, 287–337 (1985).
- Altshuler, B. L., Aronov, A. G., Gershenson, M. E. & Sharvin, Yu. V. Quantum effects in disordered metal films. *Sov. Sci. Rev. A Phys.* **9**, 223–354 (1987).
- Rosenbaum, T. F. *et al.* Metal-insulator transition in a doped semiconductor. *Phys. Rev. B* **27**, 7509–7523 (1983).
- Millis, A. J. & Lee, P. A. Spin-orbit and paramagnon effects on magnetoconductance and tunneling. *Phys. Rev. B* **30**, 6170–6173 (1984); erratum **31**, 5523–5524 (1985).
- Ishikawa, Y., Tajima, K., Bloch, D. & Roth, M. Helical spin structure in manganese silicide MnSi. *Solid State Commun.* **19**, 525–528 (1976).

Acknowledgements

We thank E. Abrahams, P.W. Adams, D.A. Browne, P. Littlewood and A. Ruckenstein for discussions. J.F.D. acknowledges the support of the Louisiana Board of Regents. J.F.D. and Z.F. acknowledge the support of the National Science Foundation.

Correspondence and requests for materials should be addressed to J. F. D. (e-mail: ditusa@phys.lsu.edu.)

Communication at millimetre–submillimetre wavelengths using a ceramic ribbon

C. Yeh*, F. Shimabukuro*†, P. Stanton*, V. Jamnejad*, W. Imbriale* & F. Manshadi*

* Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA
† The Aerospace Corp., El Segundo, California 90009, USA

Following the discovery by Kao and Hockman^{1–3} that ultra-low-loss optical fibres could be made from pure silica through the elimination of impurities, the ability to guide signals effectively at optical wavelengths has been assured. But there remains an important region of the spectrum—from 30 to 3,000 GHz (the millimetre–submillimetre band)—where low-loss waveguides are unknown. The main problem here in finding low-loss solids is no longer one of eliminating impurities, but is due to the presence of intrinsic vibration absorption bands^{4–6}. And the use of highly conducting materials is also precluded owing to high skin-depth losses^{7,8} in this part of the spectrum. Here we show that a combination of material and waveguide geometry can circumvent these difficulties. We adopt a ribbon-like structure with an aspect ratio of 10:1, fabricated from ceramic alumina (Coors' 998 Alumina), and the resulting waveguide has an attenuation factor of less than 10 dB km^{-1} in the millimetre–submillimetre band. This attenuation is more than 100 times smaller than that of a

typical ceramic (or other dielectric) circular rod waveguide and is sufficient for immediate application.

Examination of the fundamental equation^{7,8,9-11} governing the attenuation constant (α) of a dominant mode guided by a simple solid dielectric waveguide surrounded by lossless dry air shows that it is dependent on the material loss factor ($\tan \delta_1$) and the dielectric constant (ϵ_1) of the dielectric material and the geometrical size and shape of the guiding structure, described by the geometrical loss factor^{10,11} ($\epsilon_1 R$), all of which are related to the attenuation constant (in dB m⁻¹) as follows:

$$\alpha = 8.686\pi(1/\lambda_0)(\epsilon_1 R \tan \delta_1) \quad (1)$$

where λ_0 is the free-space wavelength. The geometrical loss factor is given by:

$$\epsilon_1 R = \frac{\epsilon_1 \int_A (\mathbf{E}_1 \cdot \mathbf{E}_1^*) dA}{377 [\int_A \mathbf{e}_z \cdot (\mathbf{E}_1 \times \mathbf{H}_1^*) dA + \int_{A_0} \mathbf{e}_z \cdot (\mathbf{E}_0 \times \mathbf{H}_0^*) dA]} \quad (2)$$

where $\mathbf{e}_z$ is the unit vector in the direction of propagation, A and A_0 are, respectively, the cross-sectional areas of the core and the surrounding region and $(\mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{E}_0, \mathbf{H}_0)$ are, respectively, the modal electric and magnetic field vectors of the guided mode in the core region and in the surrounding region. As the material loss factor and the dielectric constant of a solid are fixed, the only way to reduce the attenuation constant is to find the proper cross-sectional geometry of the waveguide. After performing a systematic normal-

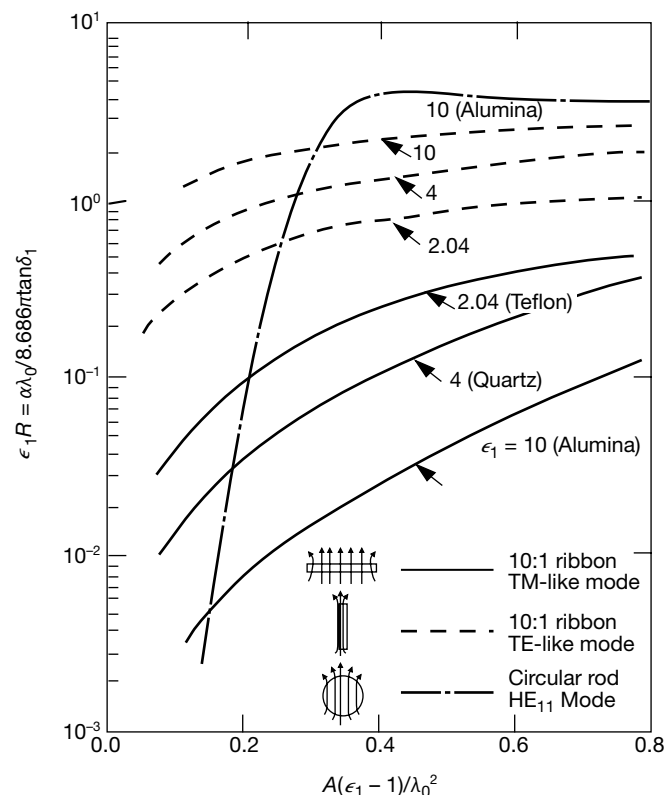


Figure 1 The geometrical loss factor $\epsilon_1 R$ as a function of the normalized cross-sectional area $A(\epsilon_1 - 1)/\lambda_0^2$. Here, A is the cross-sectional area of the waveguide, ϵ_1 is the relative dielectric constant of the dielectric guide and λ_0 is the free-space wavelength. Dielectric ribbons with an aspect ratio of 10 and an alumina circular rod are considered. For the ribbon case, the geometrical loss factors for the dominant TM-like (low-loss) and TE-like (high-loss) modes are obtained for three different dielectric materials: alumina with $\epsilon_1 = 10$, quartz with $\epsilon_1 = 4$ and teflon with $\epsilon_1 = 2.04$. The case for the alumina circular rod supporting the dominant HE_{11} mode is shown for comparison purposes. The alumina ribbon supporting the TM-like mode provides the most dramatic reduction in the geometrical loss factor when compared with that for the alumina circular rod. Sketches of transverse electric field lines for the TE-like, TM-like and HE_{11} modes are also shown.

mode study on a variety of geometries, we have shown¹²⁻¹⁴ that a ribbon-shaped guide made with low-loss, high-dielectric-constant ceramic material, such as alumina, can yield an attenuation constant for the dominant TM (transverse magnetic)-like mode of less than 0.005 dB m⁻¹. (Two dominant modes with no cut-off frequency can be supported by this ceramic ribbon structure⁹: a TE (transverse electric)-like dominant mode with most of its electric field aligned

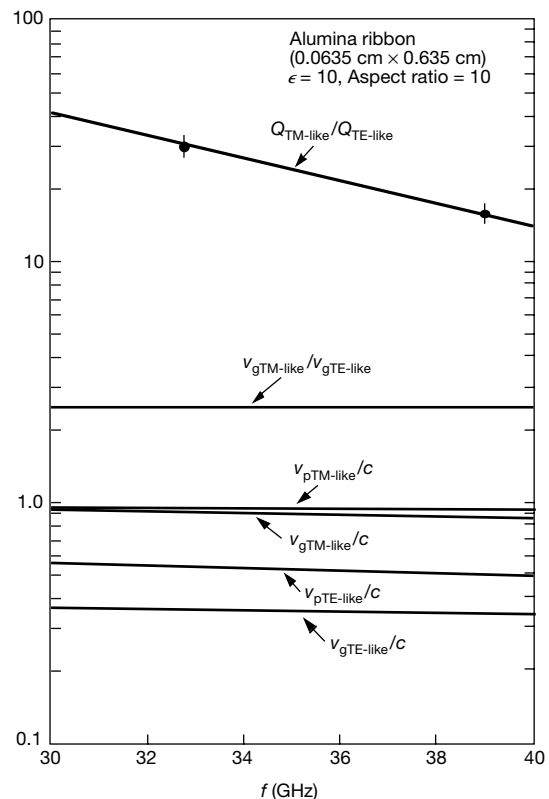


Figure 2 Ratios of $Q_{TM-like}/Q_{TE-like}$, $v_{gTM-like}/v_{gTE-like}$, $v_{pTM-like}/c$, $v_{gTM-like}/c$, $v_{pTE-like}/c$ and $v_{gTE-like}/c$ versus the frequency f (in GHz). Values taken for a 10:1 aspect ratio alumina ribbon with dimensions 0.0635 cm $\times$ 0.635 cm and dielectric constant $\epsilon_1 = 10$, supporting the TM-like mode or the TE-like mode. Here, $Q_{TM-like}$, $v_{gTM-like}$, $v_{pTM-like}$ and $Q_{TE-like}$, $v_{gTE-like}$, $v_{pTE-like}$ are the resonant Q , the group velocity and the phase velocity of the TM-like and TE-like modes, respectively. The velocity of light in vacuum is c . The ratio $Q_{TM-like}/Q_{TE-like}$ can vary from 42 at 30 GHz to 14 at 40 GHz. This means that, for a low-loss alumina ribbon with a loss tangent of 0.00005, if $Q_{TE-like}$ is measured as 22,760 at 30 GHz (this Q value is well within the measurement capability of our apparatus), then $Q_{TM-like}$ must be 955,900 (this Q value is well beyond the measurement capability of any known room-temperature resonant-cavity apparatus). To ensure that this measurement capability is not exceeded, so that $Q_{TE-like}$ and $Q_{TM-like}$ may both be measured directly in order to verify the calculated ratio ($Q_{TM-like}/Q_{TE-like}$), high-loss alumina ribbon made by coating the low-loss alumina ribbon with a thin layer of dried Indian ink is used. Measured results are shown as data points and calculated results are shown as curves. Excellent agreement is found. A photograph of the alumina-ribbon waveguide resonator is also shown.

Table 1 Measured Q , attenuation constant and loss tangent for ceramic waveguides

Batch number	Frequency (GHz)	$Q_{TE-like}$	$Q_{TM-like}$	α_{TM} (dB m ⁻¹)	$\tan\delta_1$ ($\times 10^{-4}$)
1	38.60	3,860 $\pm$ 300	64,700 $\pm$ 4,800	0.062 $\pm$ 0.005	2.8 $\pm$ 0.21
1	32.80	3,920 $\pm$ 290	121,700 $\pm$ 9,000	0.026 $\pm$ 0.002	2.8 $\pm$ 0.21
2	38.89	6,480 $\pm$ 490	103,700 $\pm$ 7,800	0.035 $\pm$ 0.003	1.59 $\pm$ 0.12
2	32.98	7,233 $\pm$ 540	216,990 $\pm$ 1,6300	0.014 $\pm$ 0.001	1.59 $\pm$ 0.12
3	39.96	10,948 $\pm$ 450	10,948 $\pm$ 450	1.45 $\pm$ 0.09	1.0 $\pm$ 0.04
3	30.03	11,117 $\pm$ 500	11,117 $\pm$ 500	1.17 $\pm$ 0.08	1.0 $\pm$ 0.04

Three batches of alumina material samples were obtained from Coors Ceramic Company (Golden, Colorado, USA). Batch 1, made from Coors' Superstrate 996 S20-71 (99.6%, Hirel, Thin Film Substrate), contains ribbons with dimensions 0.0635 cm $\times$ 0.635 cm $\times$ 11.43 cm. Batch 2, made from Coors' extruded 998 Alumina (99.8% alumina) rectangular rod, contains ribbons with dimensions 0.0635 cm $\times$ 0.635 cm $\times$ 91 cm. Batch 3 is Coors' extruded 998 Alumina (99.8% alumina) circular rod with dimensions 0.244 cm (diameter) $\times$ 91 cm (length). Numerous repeated measurements are made on these samples at various frequencies within the frequency band from 30–40 GHz, using the waveguide resonator technique. As Batch 3 contains a 91-cm-long circular alumina rod, the guiding property of the dominant mode is independent of the orientation of the transverse electric field. Thus, $Q_{TE-like} = Q_{TM-like} = Q_{HE}$ for the circular rod². Here Q_{HE} is the Q value for the HE_{11} mode on a circular dielectric rod.

parallel to the major axis of the ribbon and a TM-like dominant mode with most of its electric field aligned parallel to the minor axis of the ribbon, as sketched in Fig. 1.) As a comparison, at an operating-frequency band of around 100 GHz, the attenuation constant for the Teflon dielectric waveguide is 1.3 dB m⁻¹, for the usual metallic rectangular waveguide is 2.4 dB m⁻¹ and for the microstripline is 3 dB m⁻¹ (ref. 15). This remarkable low-loss behaviour of a ceramic ribbon guiding the dominant TM-like mode, as well as the loss behaviour of a ceramic ribbon guiding the TE-like mode, is shown in Fig. 1. A suitable choice of configuration and dielectric constant can significantly reduce the geometrical loss factor for the TM-like mode.

We draw three conclusions from our analytical investigation:

(1) Much lower geometrical loss factors are obtained for high-aspect-ratio ribbon waveguide with high dielectric constant. For example, when the normalized cross-sectional area, $A(\epsilon_1 - 1)/\lambda_0^2$, is 0.4, the geometrical loss factor (as well as the attenuation factor) for this ribbon supporting the dominant TM-like mode is about 140

times smaller than that for a circular rod with the same cross-sectional area supporting the dominant HE_{11} mode.

(2) In the low-loss region, that is, $\epsilon_1 R < 0.05$, the geometrical loss factor curve for the 10:1 ribbon with dielectric constant $\epsilon_1 = 10$ supporting the TM-like mode is much flatter than that for the circular rod supporting the HE_{11} mode. This indicates that the geometrical loss factor for the ribbon is insensitive to small deviations of the normalized cross-sectional area of the ribbon, whereas the geometrical loss factor for the circular rod is very sensitive to size changes in the rod. This means that TM-like mode on the ribbon is a very stable mode, not easily disturbed by any geometrical imperfections.

(3) Separation of the geometrical loss curves for the TE-like and TM-like modes becomes larger for larger dielectric constant of the guiding ribbon. Furthermore, there is a definite relationship between the geometrical loss curves for the two modes. These facts are very significant, because they can be used to devise a fundamentally new way to measure the super-low-loss characteristics of TM-like mode guided along ceramic ribbon.

We used the cavity resonator method¹⁶ to verify experimentally the very low attenuation constant of the ceramic ribbon waveguide. By measuring the Q of the alumina-ribbon resonator as shown in Fig. 2, we obtained the attenuation constant of the dielectric waveguide α (in dB m⁻¹) according to the formula¹⁷:

$$\alpha = 8.686(v_p/v_g)(\beta/2Q) = (8.686\pi/Q)(v_g/c)(1/\lambda_0) \quad (3)$$

where β is the propagation constant of the mode under consideration, and v_p and v_g are the phase and the group velocity, respectively, of that mode. Several factors affect the accuracy and sensitivity of this technique: the alignment of the dielectric waveguide with the coupling holes, the alignment of the parallel plates, the uniformity or the straightness of the dielectric waveguide, the coupling or radiation losses and the metallic wall losses of the plates. Previous work¹⁶ shows that the maximum Q that can be reliably measured by this technique in the K_a band (26.5–40 GHz) is approximately 30,000. This limits the smallest value of the measured attenuation constant to around 0.1 dB m⁻¹, whereas we expect the ceramic ribbon to exhibit a value around a tenth of that.

To circumvent this difficulty, we developed an indirect way to measure this ultra-low attenuation. As seen from equations (1–3), there is a definite relationship between the attenuation constants or Q values for the TM-like and TE-like modes on a dielectric ribbon:

$$r_\alpha = (\alpha_{TE-like}/\alpha_{TM-like}) = (Q_{TM-like}/Q_{TE-like})(v_{gTM-like}/v_{gTE-like}) \\ = (\epsilon_1 R)_{TE-like}/(\epsilon_1 R)_{TM-like} \quad (4)$$

where $v_{gTE-like}$ and $v_{gTM-like}$ are the group velocity of the TE-like and TM-like modes, respectively, and $Q_{TE-like}$ and $Q_{TM-like}$ are the Q values for the TE-like and TM-like modes, respectively. Measured results are displayed in Fig. 2 as data points and calculated theoretical results are given as curves. Excellent agreement between measured values and theoretical values affirms the correctness of the derived theoretical ratio ($Q_{TM-like}/Q_{TE-like}$). This relationship can

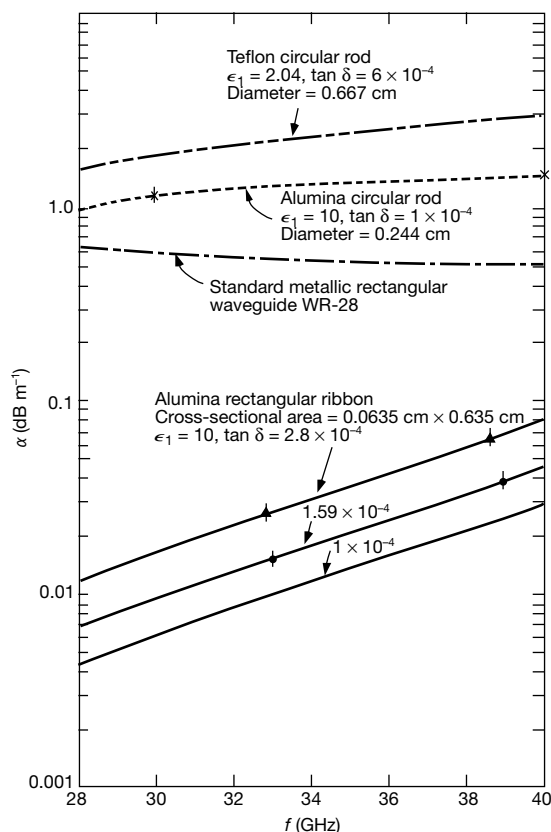


Figure 3 Attenuation constant α in dB m⁻¹ for the low-loss dominant mode in various guiding structures versus frequency f (in GHz) in the K_a band (26.5–40 GHz). Measured results are shown as data points and theoretical results are shown as curves. Excellent agreement is found.

now be used reliably to obtain $Q_{\text{TM-like}}$ when $Q_{\text{TE-like}}$ is known.

Using this technique, we measured the Q and the attenuation constant for the TM-like mode on ultra-low-loss alumina ribbon. Measured results are summarized in Table 1. The uncertainty in the measured value is discussed above. The measured data points as well as the calculated results are displayed in Fig. 3. Excellent agreement between the experimental data and theoretical results can be seen. The importance of the geometry of the guide is apparent from these results. At 30 GHz, the alumina rectangular ribbon with aspect ratio of 10 and a loss tangent of 0.000159 can support a low-loss TM-like mode with an attenuation constant at 0.0098 dB m^{-1} , which is 120 times less than that for the dominant mode on an alumina circular rod with similar cross-sectional area and a loss tangent of 0.0001. It is also 61 times less than that for the dominant mode in a standard metallic rectangular waveguide (WR-28). Significant improvement to less than 0.006 dB m^{-1} can easily be obtained for the ribbon if the same alumina material that was used for the circular rod is used.

To demonstrate the viability of the alumina ribbon as an actual transmission line, we have performed the following experiment. Shown in Fig. 4a are two horns, one transmitting and one receiving, separated by a free-space distance of 86 cm. A 120-ps pulse is emitted from the transmitting horn and is received by the receiving horn as pulse B after traversing through this free-space distance of 86 cm. We performed another experiment, where a 91-cm-long ceramic-ribbon waveguide with cross-sectional area of $0.0635 \text{ cm} \times 0.635 \text{ cm}$ was inserted between the horns. Specially designed transitions are used to maximize launching and receiving efficiencies. A launching (or receiving) efficiency of 84% (or loss of less than 0.825 dB) at 39.86 GHz has been measured for an exponential launching horn. The same 120-ps pulse is now sent through this ceramic-ribbon waveguide structure. The received pulse is labelled pulse A. The received pulses for these two cases are displayed in Fig. 4a. The received signal through the ceramic-ribbon waveguide is at least 21 dB greater than that through free space, providing clear evidence that the signal can easily be guided by the ceramic ribbon.

The ceramic-ribbon waveguide is an open structure surrounded by dry air. How to support such a structure is an important consideration. One supporting structure appears to be most promising—support made with plastic fishing-line, that is, nylon thread. (See upper photograph in Fig. 4.) Thin nylon (low dielectric constant) threads, spaced at least 10 cm apart, are strung across wooden rails separated by 5 cm (far enough apart so that the exterior guiding field at the ribbon edges has decayed to negligible value). Ceramic-ribbon waveguide can simply be laid on top of the nylon threads along the middle of the rails. The nylon threads can easily support the ceramic ribbon. Any perturbation caused by the nylon-thread support on the propagation characteristics of the guided TM-like mode on the ceramic ribbon waveguide is not detectable.

Another practical problem is how to join sections of ceramic-ribbon waveguides. We have discovered that a 'shiplap' joint may be used to provide a strong bond between two ends of ceramic ribbon. A picture of the shiplap joint is shown in Fig. 4. A quarter-inch length of the jointing ribbon end is ground to a thickness of 0.0317 cm, which is half the original thickness of the ribbon. The jointing end of the other ceramic ribbon is prepared similarly. The ends are lapped together, aligned and then glued with 'super glue', resulting in a strong bond. We found no measurable loss attributable to the joint. Finally, to further understand how power is being guided along a high-dielectric-constant ($\epsilon_1 = 10$) thin ribbon structure that can provide the very low attenuation constant for the dominant mode, see Fig. 4b.

Just as the first 20-dB km^{-1} optical fibre made in the late 1960s produced a revolution in optical communication¹⁻³, so may the attainment of 10 dB km^{-1} ceramic ribbon provide an opening to communications at 30–3000 GHz. Some immediate applications in this frequency band may be: transporting signals for Jet Propulsion

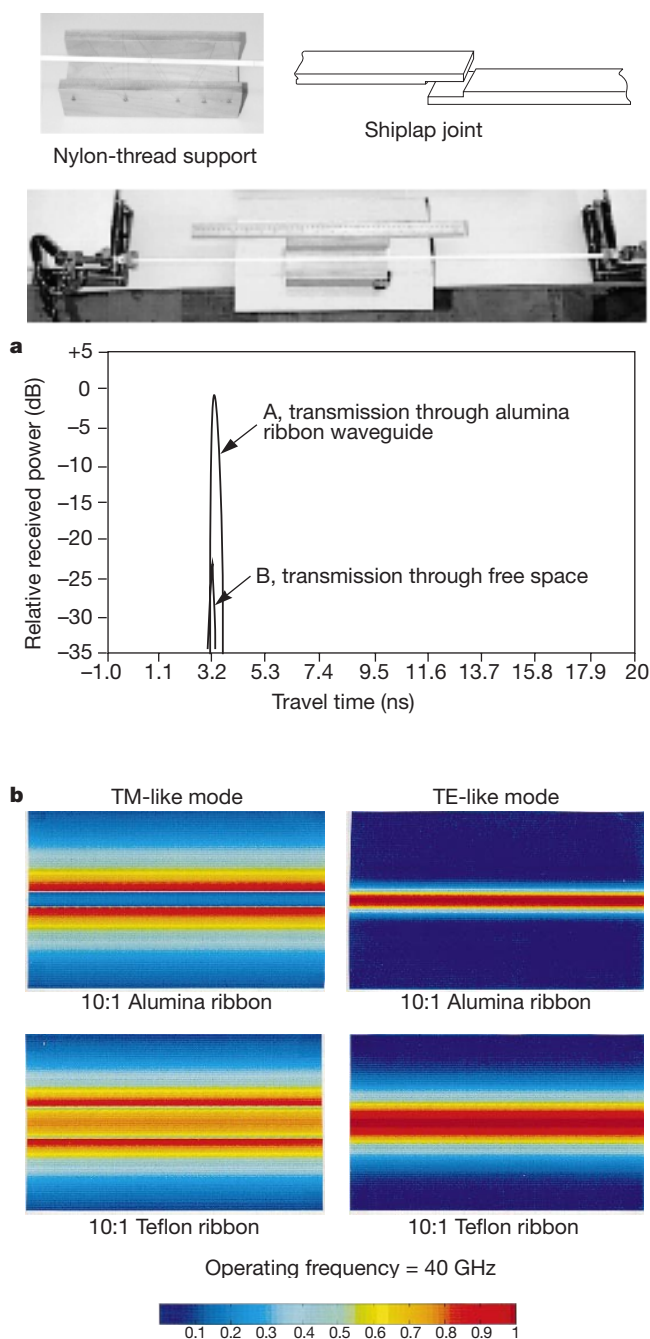


Figure 4 Transmission by ceramic ribbon and its power distribution. **a**, Comparison between the received pulses for a 120-ps pulse sent through horns linked by the alumina-ribbon waveguide (pulse A) and that linked by free-space (pulse B). The picture of a 91-cm-long alumina-ribbon waveguide with cross-sectional area of $0.0635 \text{ cm} \times 0.635 \text{ cm}$ with launching and receiving horns is shown. The slight delay of the arrival of pulse A indicates that, owing to wave guidance by the alumina ribbon, pulse A is being guided by the ribbon and is propagating at the group velocity of the TM-like mode on this structure. This guided group velocity is slower than c , the free-space group velocity. To prevent sagging, the ceramic waveguide is supported by taut, widely spaced nylon threads. Inset, to make a long distance transmission line, sections of ceramic ribbon may be joined together in a 'shiplap' manner with 'super glue'. No measurable loss was noted due to the nylon-thread support or the joins. **b**, Normalized power-intensity distribution for the dominant normal modes (TM-like mode and TE-like mode) on a 10:1 dielectric ribbon structure. The highest power intensity is indicated in red and the lowest power intensity is indicated in blue. The cross-sectional sizes are chosen for single-mode operation at 40 GHz. Unlike the traditional dielectric waveguide case, the distinguishing feature here is that there is a dip in the power intensity for the TM-like mode within the thin alumina guiding ribbon.

Laboratory's deep space network and on spacecraft, transporting signals for detectors, sensors and phase array antennas or carrying long distance signals, to name a few. It is conceivable that ceramic-ribbon waveguide may be the backbone of future communication systems in this frequency band. □

Received 3 June 1999; accepted 31 January 2000.

1. Kao, K. C. & Hockman, G. A. Dielectric fiber surface waveguides for optical frequencies. *Proc. IEEE* **133**, 1151–1158 (1966).
2. Agrawal, G. P. *Fiber Optic Communication Systems* (Wiley Series in Microwave and Optical Engineering, New York, 1997).
3. Marcuse, D. *Light Transmission Optics* (Van Nostrand-Reinhold, New York, 1972).
4. Afsar, M. N. & Button, K. J. Millimeter-wave dielectric measurement of materials. *Proc. IEEE* **73**, 131–153 (1985).
5. Birch, R., Dromey, J. D. & Lisurf, J. The optical constants of some common low-loss polymers between 4 and 40 cm^{-1} . *Infrared Phys.* **21**, 225–228 (1981).
6. Afsar, M. N. Precision dielectric measurements of nonpolar polymers in millimeter wavelength range. *IEEE Trans. Microwave Theor. Tech.* **33**, 1410–1415 (1985).
7. Yeh, C. in *American Institute of Physics Handbook* (ed. Gray, D. E.) 3rd edn (McGraw Hill, New York, 1972).
8. Ramo, S., Whinnery, J. R. & Van Duzer, T. *Fields and Waves in Communication Electronics* 2nd edn (Wiley, New York, 1984).
9. Yeh, C. Elliptical dielectric waveguides. *J. Appl. Phys.* **33**, 3235–3243 (1962).
10. Yeh, C. Attenuation in a dielectric elliptical cylinder. *IEEE Trans. Antenna Propag.* **11**, 177–184 (1963).
11. Yeh, C., Shimabukuro, F. I. & Chu, J. Ultra-low-loss dielectric ribbon waveguide for millimeter/submillimeter waves. *Appl. Phys. Lett.* **54**, 1183–1185 (1989).
12. Yeh, C., Ha, K., Dong, S. B. & Brown, W. P. Single-mode optical waveguides. *Appl. Opt.* **18**, 1490–1504 (1979).
13. Taflov, A. *Computational Electrodynamics, the Finite-Difference Time-Domain Method* (Artech House, Norwood, MA, 1995).
14. Yeh, C., Casperson, L. & Szejn, B. Propagation of truncated gaussian beams in multimode fiber guides. *J. Opt. Soc. Am.* **68**, 989–993 (1978).
15. Koul, S. K. *Millimeter Wave and Optical Dielectric Integrated Guides and Circuits* (Wiley Series in Microwave and Optical Engineering, New York, 1997).
16. Shimabukuro, F. I. & Yeh, C. Attenuation measurement of very low loss dielectric waveguides by the cavity resonator method applicable in the millimeter/submillimeter wavelength range. *IEEE Trans. Microwave Theor. Tech.* **36**, 1160–1166 (1988).
17. Yeh, C. A relation between α and Q . *Proc. Inst. Radio Eng.* **50**, 2145 (1962).

Acknowledgements

We thank C. Stelzried, A. Bhanji, D. Rascoe and M. Gatti of JPL for their encouragement and support. Expert assistance from M. Ostrander in machining, R. Cirillo in experimental measurements and C. Copeland in graphic works is appreciated. The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

Correspondence and requests for materials should be addressed to C. Y. (e-mail: cavour.yeh@jpl.nasa.gov).

Functional hydrogel structures for autonomous flow control inside microfluidic channels

David J. Beebe[†], Jeffrey S. Moore^{*}, Joseph M. Bauer^{*}, Qing Yu^{*}, Robin H. Liu^{*}, Chelladurai Devadoss^{*} & Byung-Ho Jo^{*}

^{*} The Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[†] Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

Hydrogels have been developed to respond to a wide variety of stimuli^{1–6}, but their use in macroscopic systems has been hindered by slow response times (diffusion being the rate-limiting factor governing the swelling process). However, there are many natural examples of chemically driven actuation that rely on short diffusion paths to produce a rapid response⁷. It is therefore expected that scaling down hydrogel objects to the micrometre scale should greatly improve response times. At these scales, stimuli-responsive

hydrogels could enhance the capabilities of microfluidic systems by allowing self-regulated flow control. Here we report the fabrication of active hydrogel components inside microchannels via direct photopatterning of a liquid phase. Our approach greatly simplifies system construction and assembly as the functional components are fabricated *in situ*, and the stimuli-responsive hydrogel components perform both sensing and actuation functions. We demonstrate significantly improved response times (less than 10 seconds) in hydrogel valves capable of autonomous control of local flow.

Conventional microactuators (using, for example, electromagnetic, electrostatic or thermopneumatic effects) require external power for operation and relatively complex assembly, which limits their use in practical systems⁸. Stimuli-responsive hydrogels have a significant advantage over conventional microfluidic actuators owing to their ability to undergo abrupt volume changes in response to the surrounding environment without the requirement of an external power source. Difficulties in integrating conventional micrometre-scale components into functional systems using traditional approaches^{8,9} have limited the potential benefits of emerging microfluidic systems. Unconventional approaches are needed to overcome these difficulties in microscale integration. Recently, several groups have explored new fabrication methods that show promise^{10–14}. Our approach combines lithography, photopolymerization and microfluidics to create functional components (valves) within microchannels for local flow control. We believe that this approach has widespread applications for flow regulation in microfluidic systems.

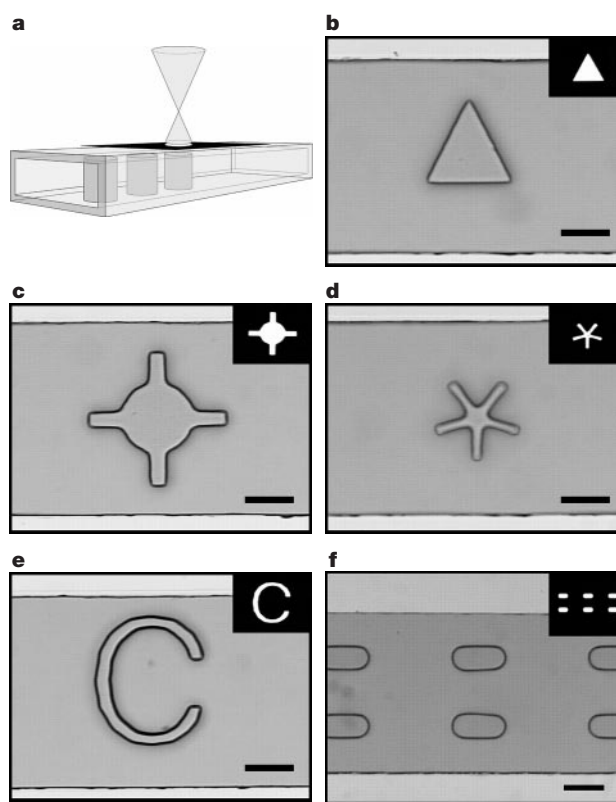


Figure 1 A diagram of the fabrication method and images demonstrating a variety of shapes that were polymerized within 35 seconds. **a**, The fabrication method. **b**, A polymerized hydrogel demonstrating the ability to pattern high-definition straight edges. The corresponding photomask is shown at a reduced size in the upper right corner of each picture. **c**, **d**, Structures illustrating the generation of convex and concave surfaces. **e**, A structure with high-aspect-ratio features. Imperfections in the mask were transferred to the structure, further demonstrating the high fidelity of the photolithographic process. **f**, The simultaneous polymerization of multiple structures with a single exposure of ultraviolet light. Scale bars: **b–e**, 250 μm ; **f**, 500 μm .

The general approach we use is to flow a mixture of monomers and a photoinitiator into the microchannel, and irradiate the mixture through a photomask (Fig. 1a). In a typical procedure, transparent channels ranging from 500 to 2,000 μm wide and 50 to 180 μm deep are filled with a photopolymerizable liquid consisting of acrylic acid and 2-hydroxyethyl methacrylate (in a 1:4 molar ratio), ethylene glycol dimethacrylate (1 wt%) and a photoinitiator (3 wt%). The liquid is allowed to reach a quiescent state and is then exposed to ultraviolet light through a photomask placed on top of the channel. Polymerization times vary depending on light intensity, photoinitiator and monomer mixture, and can be less than 20 seconds using Irgacure 651 as the photoinitiator and the filtered light source from a standard fluorescence microscope. When the polymerization is finished, the channel is flushed with water to remove the unpolymerized liquid. This method allows pH-responsive hydrogels of different shapes and sizes to be integrated directly into microfluidic systems.

As seen in Fig. 1, the pattern of the photomask is transferred to the polymerized object with high fidelity. The minimum feature size is 25 μm , coinciding with the minimum resolution of the photomask. Confocal micrographs of an object polymerized in a 180- μm -deep channel by irradiation through a circular mask reveal a slightly tapered cylindrical shape, with a smaller diameter at the bottom than the top. The polymerized hydrogel is in contact with both the top and bottom of the channel. The fabrication of multiple structures can be performed either sequentially by moving the mask, or simultaneously by using a mask with a multi-structure pattern (Fig. 1f). Furthermore, multiple hydrogels of different chemical compositions can be fabricated in a single channel by sequential fabrication, as demonstrated below for the flow sorter.

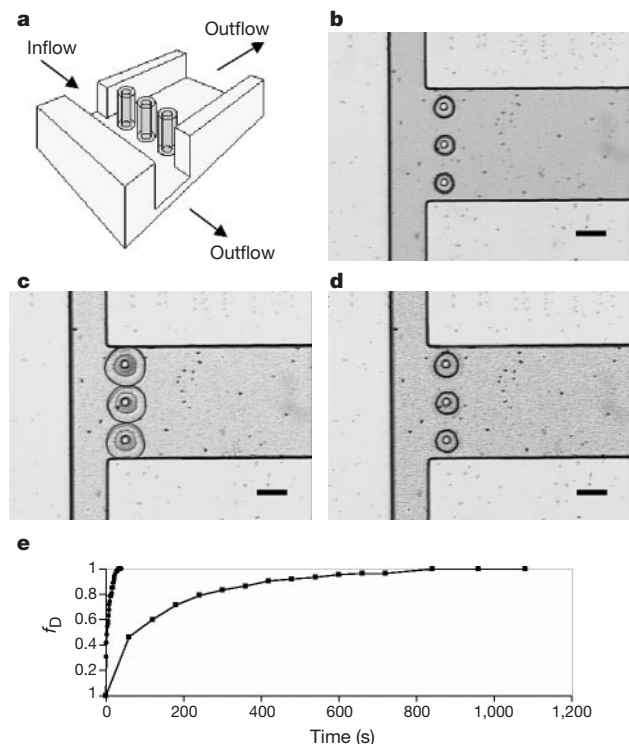


Figure 2 Prefabricated posts in a microchannel serve as supports for the hydrogels, improving stability during volume changes. **a**, A diagram of the hydrogel jackets around the posts. **b**, The actual device after polymerization of the hydrogel. **c**, The hydrogel jackets block the side channel branch in their expanded state. **d**, The contracted hydrogels allow fluid to flow down the side branch. **e**, The improvement in time response of the hydrogel jacket design (circles) versus an alternative design that uses a single larger cylindrical structure in the same size channel (squares). f_D is the fractional change in diameter. Scale bars, 300 μm .

The time response of the volume change approximately follows the square of the dimension as the hydrogel objects reversibly expand and contract, depending on the pH of the surrounding environment. We characterize this dynamic behaviour by measuring the step response of the hydrogel expansion. A step response of less than 10 seconds was observed for a 100- μm -diameter cylindrical structure in a 50- μm -deep channel, but this configuration was mechanically unstable. Hydrogel objects tend to buckle or migrate during a volume change if their lateral dimensions are smaller than the channel height. To fabricate stable objects with fast response times, we polymerize the hydrogel structures around prefabricated posts. The posts provide a robust support, and also improve time response owing to the short diffusion path of the hydrogel jackets surrounding the posts. An array of hydrogel-coated posts can control the flow in large channels, as shown in Fig. 2. The step response for expansion of this array valve design is 8 seconds (the contraction step response is of the same order). In contrast, an alternative valve design that uses a single larger cylindrical structure in the same size channel has a step response of 130 seconds over the same pH range. In this case, the post design accelerates the response time by a factor of 16 (Fig. 2e). This integration of hydrogels into microfluidic systems provides the scaling necessary to overcome the primary drawback (slow time response) of hydrogels.

Expansion and contraction of the hydrogels can regulate the fluid flow in microchannels. Measuring the pressure drop at constant flow rate over a channel containing hydrogel structures reveals this behaviour. To illustrate this, we use a flow rate of 0.15 ml min^{-1} to pump solutions of various pH values past oval-shaped structures (300 $\times$ 700 μm , contracted state) polymerized in a 1 $\times$ 9 array along the length of a 1,000- μm -wide glass channel. In an acidic

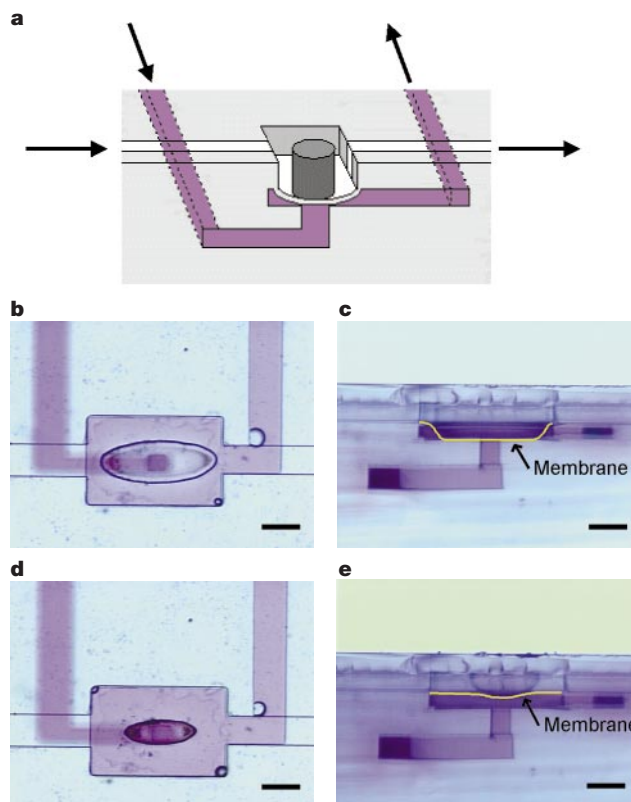


Figure 3 A shut-off valve. **a**, A diagram of the valve design. The arrows denote the direction of fluid flow. **b**, **c**, The hydrogel structure expands and deforms a membrane, blocking flow in an adjacent channel. The images on the left show a top view of the device, and the images on the right show the side view. The fluid in the blocked channel has been dyed for visualization purposes. **d**, **e**, The hydrogel contracts, and the membrane returns to a position that allows flow in the adjacent channel. In **c** and **e**, the boundary of the membrane has been outlined for clarity (yellow). Scale bars, 250 μm .

environment, the objects are in a contracted state and produce a pressure drop of 0.09 pounds per square inch (p.s.i.). Upon raising the pH above the transition point of these hydrogels, the objects fully expand, causing the pressure drop to increase almost eight-fold. This simple-to-fabricate device thus functions as a pH-sensitive throttle valve for microfluidic systems.

Through appropriate design, a single hydrogel component can sense the chemical environment in one channel and regulate the flow in an adjacent channel, as shown in Fig. 3. This device contains a flexible membrane that can deform to block the flow in an adjacent channel. The hydrogel structure polymerized in the channel above the membrane expands or contracts as the surrounding pH is changed. The force associated with these volumetric changes is sufficient to deform the membrane and consequently control the flow in the lower channel. It is easy to imagine extension of this demonstration to antigen-responsive hydrogels⁶ that could serve as devices in self-regulated drug delivery or biosensors.

An additional demonstration of the versatility of this approach is the fabrication of a self-regulated 'flow sorter'. This device consists of a 'T' channel in which the entrance to each branch is gated with a hydrogel structure of unique chemical composition. The hydrogel for one branch expands at high pH and contracts at low pH, while a hydrogel of a different composition gates the other branch and exhibits an inverse behaviour (that is, contracts at high pH and expands at low pH). The device and a graph of its resulting output responses are shown in Fig. 4. This device automatically directs the flow in the centre channel down one branch or the other depending on the pH. In a certain pH range (5.7–6.8), both hydrogel valves swell to seal the channel. Each hydrogel valve performs the sensing, actuating and regulating functions normally performed by discrete components (valve, sensors, electronics) in a traditional system. By tailoring the chemical composition of the hydrogel, the output response (the slope and position of the volume transition) can be modified, allowing pH-sensitive hydrogels to be used in a variety of applications.

The ability to fabricate functional structures within microfluidic channels has the potential to simplify greatly the processes required to build complex microfluidic systems. Our approach eliminates difficult microscale assembly and the need for electronics for

sensing and actuation. In addition, fast response times are achieved owing to the short diffusion paths in microchannels. As noted above, the approach is not limited to pH-response hydrogels, thus enabling a wide range of functional components. The microfluidic/photopolymerization fabrication method described here provides an approach that could be extended to build multifunctional microfluidic systems, allowing complex fluidic processes to be performed autonomously. □

Methods

Photopolymerization

The photomasks were prepared by printing patterns on transparency films using a high-resolution commercial printer system (Linotype Herkules Imagesetter) with a resolution of 5,080 dots per inch. The photoinitiator, Irgacure 651, is the registered trade name of 2,2-dimethoxy-2-phenyl acetophenone (Ciba Specialty Chemicals). A near-ultraviolet filter cube (U-MNUA, type BP360-370) was used on an Olympus Epi-Fluorescent microscope (BX-60) to provide the energy necessary for polymerization. The band pass of the cube was 360–370 nm. The resulting hydrogel objects had side walls that deviated from the surface normal by ~6°. All images are unprocessed. The purple hue in Fig. 3 is produced by a dye added for visualization purposes. Arrows and lines have been added in Fig. 3 to highlight the edge of the membrane to clarify its location.

Device fabrication and characterization

The channels in Fig. 1 were constructed by bonding two no. 1 coverslips to a glass substrate with a UV-curable adhesive. The coverslips were placed so that their edges were parallel and separated by the desired channel width. A third coverslip served as the channel top, producing channels that were ~180 µm deep. The channel and posts for the device shown in Fig. 2 were fabricated by etching a 200-µm-thick photosensitive epoxy layer (Nano XP SU-8, Microchem Corp.) spin-coated on a Pyrex substrate. Transparent adhesive tape served as the top of the channel. Step response is the time required for the hydrogel to reach (1 - 1/e) (63.2%) of its total volume change. The device in Fig. 3 was constructed using a fabrication technique previously developed¹⁵. The device shown in Fig. 4 has no. 1 coverslips as top and bottom, with a 200-µm-thick poly(dimethylsiloxane) gasket in between to define the channel height. The hydrogel on the left was prepared from 2-(dimethylamino)ethyl methacrylate and 2-hydroxyethyl methacrylate (in a 1:4 molar ratio), ethylene glycol dimethacrylate (1.4 wt%) and Irgacure 651 as the photoinitiator (3 wt%). The hydrogel on the right was prepared using the components listed in the text. The pressure measurements were taken using a Validyne model DP-15 pressure transducer with a 0–1 p.s.i. nickel plated diaphragm. The pressure drop measurements were taken over a channel containing hydrogel ovals with their major axes parallel to the channel walls, similar to the configuration shown in Fig. 1f. To verify fluid flow in the various device designs, a variety of flow visualization techniques such as particle tracking, dyes and air bubbles were used.

Received 17 December 1999; accepted 16 February 2000.

1. Tanaka, T. *et al.* Phase transitions in ionic gels. *Phys. Rev. Lett.* **45**, 1636–1639 (1980).
2. Hu, Z., Zhang, X. & Li, Y. Synthesis and application of modulated polymer gels. *Science* **269**, 525–527 (1995).
3. Tanaka, T., Nishio, I., Sun, S.-T. & Ueno-Nishio, S. Collapse of gels in an electric field. *Science* **218**, 467–469 (1982).
4. Suzuki, A. & Tanaka, T. Phase transition in polymer gels induced by visible light. *Nature* **346**, 345–347 (1990).
5. Kataoka, K., Miyazaki, H., Bunya, M., Okano, T. & Sakurai, Y. Totally synthetic polymer gels responding to external glucose concentration: their preparation and application to on-off regulation of insulin release. *J. Am. Chem. Soc.* **120**, 12694–12695 (1998).
6. Miyata, T., Asami, N. & Uragami, T. A reversibly antigen-responsive hydrogel. *Nature* **399**, 766–769 (1999).
7. Schmidt-Nielsen, K. *Animal Physiology* (Cambridge Univ. Press, Cambridge, 1975).
8. Kovacs, G. T. A. *Micromachined Transducers Sourcebook* (WCB McGraw-Hill, Boston, 1998).
9. Trimmer, W. S. N. Microrobots and micromechanical systems. *Sensors Actuators* **19**, 267–287 (1988).
10. Smela, E., Inganäs, O. & Lundström, I. Controlled folding of micrometer-size structures. *Science* **268**, 1735–1738 (1995).
11. Breen, T., Tien, J., Oliver, S. R. J., Hadzic, T. & Whitesides, G. M. Design and self-assembly of open, regular, 3D mesostructures. *Science* **284**, 948–951 (1999).
12. Jackman, R. J., Brittain, S. T., Adams, A., Prentiss, M. G. & Whitesides, G. M. Design and fabrication of topologically complex, three-dimensional microstructures. *Science* **280**, 2089–2091 (1998).
13. Kenis, P. J., Ismagilov, R. F. & Whitesides, G. M. Microfabrication inside capillaries using multiphase laminar flow patterning. *Science* **285**, 83–85 (1999).
14. Cumpston, B. H. *et al.* Two-photon polymerization initiators for three-dimensional optical data storage and microfabrication. *Nature* **398**, 51–54 (1999).
15. Jo, B.-H., Lerberghe, L. M. V., Motsegood, K. M. & Beebe, D. J. Three-dimensional micro-channel fabrication in polydimethylsiloxane (PDMS) elastomer. *J. Micromech. Syst.* **9**, 76–81 (1999).

Acknowledgements

This work was supported by DARPA-MTO.

Correspondence and requests for materials should be addressed to D.J.B. (e-mail: dbecbe@engr.wisc.edu).

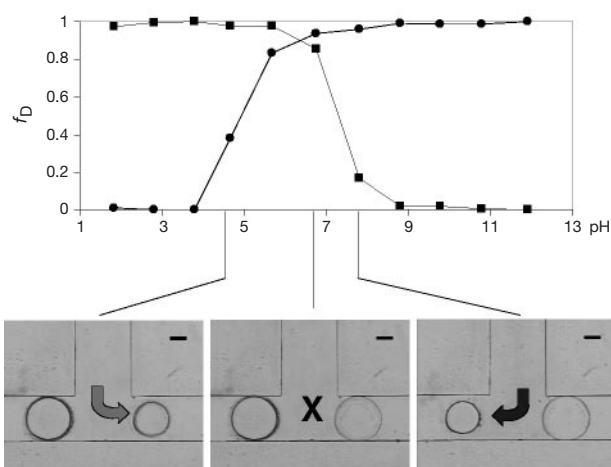


Figure 4 The volume response of two different hydrogels with respect to the pH of the surrounding fluid. Top, the fractional change in diameter (f_d) of the hydrogels with respect to pH. Bottom, images showing a device that directs ('sorts') a fluid stream on the basis of its pH. The hydrogel gating the right branch (circles) expands in base and contracts in acid. The hydrogel gating the left branch (squares) behaves in the opposite manner (expands in acid and contracts in base). The fluid enters from the centre channel at a rate of 0.05 ml min⁻¹. At a pH of 7.8, the flow is directed down the left branch. At a pH of 4.7, the flow is directed down the right branch. Both hydrogels expand to shut off the flow when the pH is changed to 6.7. Scale bars, 300 µm.

Substantial contribution to sea-level rise during the last interglacial from the Greenland ice sheet

Kurt M. Cuffey* & Shawn J. Marshall†‡

* Department of Geography, University of California, Berkeley, California 94720-4740, USA

† Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

‡ Present address: Department of Geography, University of Calgary, ES 356, 2500 University Drive, NW, Calgary, Alberta T2N 1N4, Canada

During the last interglacial period (the Eemian), global sea level was at least three metres, and probably more than five metres, higher than at present^{1,2}. Complete melting of either the West Antarctic ice sheet or the Greenland ice sheet would today raise sea levels by 6–7 metres. But the high sea levels during the last interglacial period have been proposed to result mainly from disintegration of the West Antarctic ice sheet³, with model studies attributing only 1–2 m of sea-level rise to meltwater from Greenland^{4,5}. This result was considered consistent with ice core evidence⁴, although earlier work had suggested a much reduced Greenland ice sheet during the last interglacial period⁶. Here we reconsider the Eemian evolution of the Greenland ice sheet by combining numerical modelling with insights obtained from recent central Greenland ice-core analyses. Our results suggest that the Greenland ice sheet was considerably smaller and steeper during the Eemian, and plausibly contributed 4–5.5 m to the sea-level highstand during that period. We conclude that the high sea level during the last interglacial period most probably included a large contribution from Greenland meltwater and therefore should not be interpreted as evidence for a significant reduction of the West Antarctic ice sheet.

The substantial impact that a 5-m sea-level rise would have on coastal regions provides strong motivation for understanding the origins of the Eemian historical highstand. The peculiar dynamics^{7,8} of the West Antarctic ice sheet (WAIS) (flow dominated by basal processes and grounding below sea level) may have, in theory, enabled it to rapidly disintegrate in response to a minor and transient sea-level forcing³. In contrast, it is accepted that complete melting of the Greenland ice sheet (GIS) would have required a sustained climatic forcing⁹. A substantial portion of the mass loss from the GIS occurs by surface melting in the low-elevation zone around its margin, and a modest climatic warming could rapidly

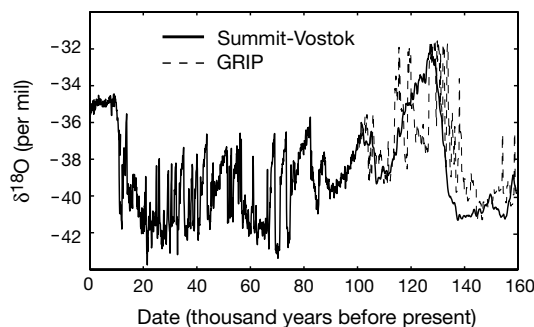


Figure 1 Assumed isotopic history. The solid line is our 'standard case' isotopic history used as model forcing. It was fabricated for use in the present study and should not be mistaken for a Greenland climate reconstruction. For comparison, the dashed line shows the GRIP record²² which has been used in recent modelling studies⁵, and which we do not use because the chronology is known to be disturbed by folding of the ice²³.

expand this zone of net melt across the gently sloping ice-sheet surface. The GIS is thus recognized to be moderately vulnerable to climatic warming, and the warming necessary to sustain destruction of the entire ice sheet is estimated to be 5–8 °C (ref. 10).

Here we argue that the current 1–2 m estimate^{4,5} for the Eemian GIS sea-level contribution is too small and that a contribution more than twice this large is plausible. The difference results primarily from a quantitative re-evaluation of the isotopic palaeothermometer ($\delta^{18}\text{O}$) used to construct the climate histories driving the evolution of the model ice sheets.

Generally, a $\delta^{18}\text{O}$ history measured from ice core samples is used to construct a climatic temperature history by assuming proportionality between deviations of mean annual temperature (ΔT) and isotopic content ($\Delta\delta^{18}\text{O}$) from modern values, so that $\Delta\delta^{18}\text{O} = \alpha\Delta T$. The earlier modelling studies assumed, as was standard practice, that the sensitivity of isotopic content to temperature $\alpha = d\delta^{18}\text{O}/dT \approx 0.67$ per mil °C⁻¹, which is the modern value for α evaluated as a spatial derivative on the ice-sheet surface. Recently, palaeotemperature reconstructions based on temperature–depth distribution^{11,12} and on gas-phase isotopic composition¹³ have revealed that the temporal isotopic sensitivity (α evaluated at one location through time) is much smaller and was, for example, approximately 0.4 per mil °C⁻¹ for the last glacial–interglacial transition in central Greenland, when corrected for changes in ocean water composition. Several reasons for the lower sensitivity have been recognized, including changes in the seasonal timing of precipitation¹⁴ and covariation of low- and high-latitude temperature changes¹⁵.

These factors are also expected to cause low α for millennial-scale climate changes within interglacial periods. For example, changes in seasonal timing of precipitation are modelled to occur in correlation with temperature changes in the Holocene epoch, because both

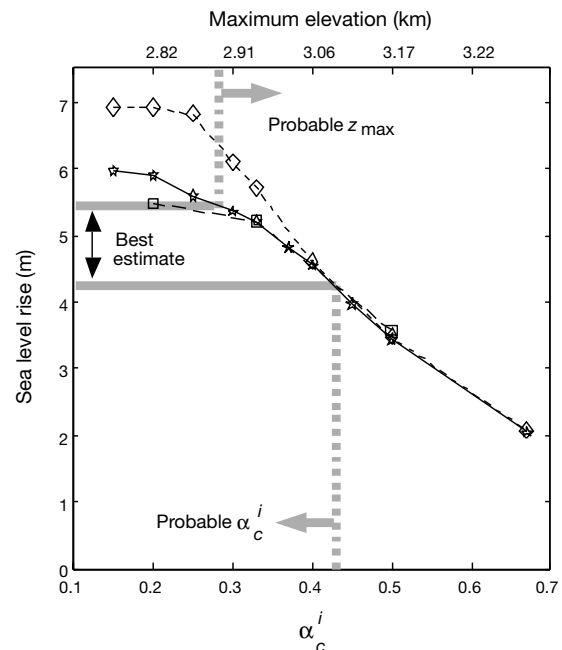


Figure 2 Calculated Eemian sea-level rise due to Greenland melting. Black stars show the maximum Eemian sea-level rise, as a function of the isotopic sensitivity α_c , using the 'standard case' isotopic forcing. The corresponding summit elevations are shown on the top axis. Diamonds show results of the same calculations, but with the isotopic lapse rate correction (see Methods) removed, to illustrate its importance. Squares represent calculations applying $\alpha_c = 0.33$ to all the pre-Eemian forcing with $\delta \leq -35$ per mil (which implies that the penultimate glacial climate was very similar to that of the last glacial). The black stars were calculated with $\alpha_c = \alpha_c^i$, which is physically less plausible but which allows direct comparison to earlier modelling work in the limit $\alpha_c^i = 0.67$, and gives a minimum lower bound for the 'best estimate' sea-level contribution.

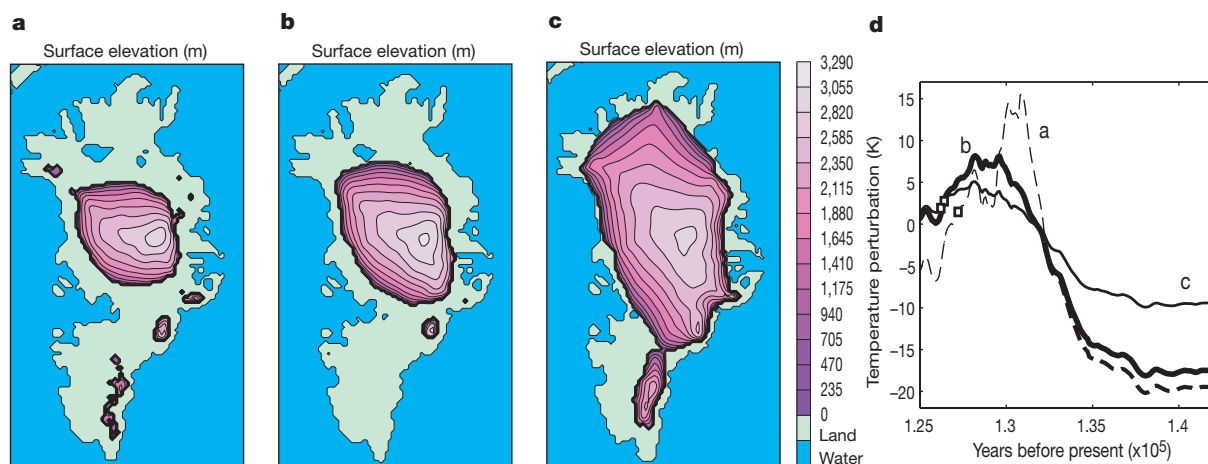


Figure 3 Calculated surface elevation maps for the Greenland ice sheet at its Eemian minimum. **a–c**, Maps of three different values for α_c : $\alpha_c = 0.2$ (with pre-Eemian $\alpha_c = 0.33$) (**a**), $\alpha_c = 0.4$ (with pre-Eemian $\alpha_c = 0.4$) (**b**), $\alpha_c = 0.67$ (with pre-Eemian $\alpha_c = 0.67$) (**c**). Map **b** is preferred. Map **c** is entirely consistent with previous modelling efforts^{4,5}. **d**, The constant-elevation climatic temperature perturbations (relative to modern

temperature) generated by these three model calculations, with the squares indicating time of minimum ice sheet volume: thick line (map **b**), thin line (map **c**), dashed lines (map **a**). The thick portion of the dashed line is also the pre-Eemian temperature history used in the three calculations shown by squares in Fig. 2.

depend on the strength of the North Atlantic thermohaline circulation¹⁶, which itself responds to variations of the Atlantic basin moisture balance¹⁷. Thus an unlikely constancy of numerous climate variables would be required in order to have α equal to its modern spatial gradient.

Indeed, a low sensitivity, α_c , for millennial-scale climate changes within interglacial periods is implied by Holocene climate history, as recorded in borehole temperatures and the Greenland cores' isotopic records. Central Greenland has cooled over the last 5 kyr, by 1.5–2.5 °C according to several independent borehole temperature analyses^{12,11,18} and this cooling is recorded as a 0.375 per mil drop in average $\delta^{18}\text{O}$ (refs 19, 20). Thus the observed α has been $\alpha_{\text{obs}} = \Delta\delta_{\text{obs}}/\Delta T_{\text{obs}} \approx 0.15$ to 0.25 per mil °C⁻¹. This indicates α_c is in the range 0.15 to 0.43 after accounting for possible elevation changes (see Methods), and therefore is similar to α for the glacial–interglacial transition, or possibly lower.

Such low values for α imply that the Eemian climate in Greenland was warmer than previously thought, and the consequent melting of the GIS more significant. To investigate this possibility, we model the evolution of the GIS in the changing climate of the last 160,000 years, using a three-dimensional coupled-ice-and-heat-flow model²¹, developed using ref. 9. The model includes elevation changes due to isostatic response of the crust, and has a horizontal grid of resolution 20 km by 20 km.

The climatic forcing for the model consists of prescriptions for temperature, accumulation rate and ablation rate at all grid points on the ice-sheet surface, as functions of time. These forcings are all governed by a history of temperature change at constant elevation $\Delta T_c(t)$ (relative to modern), using standard relationships (see Methods) incorporating surface elevation and slope, and latitude–longitude. The ΔT_c spans 160 kyr, representing the cold climate of the penultimate glacial period, the warming to the Eemian interglacial (110–130 kyr ago), and the return to glacial conditions followed by warming to the Holocene. The model is initialized by first calculating a steady-state ice sheet for modern climate, and then imposing 200 kyr of constant climate 6 °C colder than modern. Memory of this initial state is largely erased within the first 15 kyr of the model calculation; results for the Eemian are insensitive to plausible changes in the initialization temperature.

Our interest here is the Eemian fate of the ice sheet, but we run the model for another 100 kyr to the present to show that the modern GIS is accurately reproduced. The model ice sheet accurately

reproduces all gross features of GIS geography, including the volume (to within 1%), and the summit location (to within 50 km in the horizontal and 40 m in the vertical). For the most recent 98 kyr we use the ΔT_c inferred from the borehole temperature calibration^{11,18} of the GISP2 project ice core $\delta^{18}\text{O}$ record¹⁹.

Earlier forcings must be treated differently. There are at present no reliable palaeoclimate histories for the Eemian or earlier periods from any of the Northern Hemisphere ice caps. We circumvent this difficulty by fabricating a plausible climate history (meaning a history of isotopic content) and performing appropriate sensitivity tests. The crucial constraint on the history is the maximum $\delta^{18}\text{O}$ value for Eemian ice in the Summit cores, which is heavier than modern accumulation by an amount $\Delta\delta_E \approx 3.3$ per mil (ref. 22). This value fixes the maximum Eemian $\delta^{18}\text{O}$ in our fabricated history. Although the chronology of Eemian ice in the Summit cores is uncertain, there is no reason to think the $\delta^{18}\text{O}$ composition has been altered, and the combined CH_4 and $\delta^{18}\text{O}$ of O_2 gas compositions show conclusively that this ice accumulated during the Eemian²³.

Furthermore, we know the approximate duration of the Eemian as a whole based on a globally significant variety of independent but mutually consistent climate proxy records, including the Vostok ice core deuterium-isotope history which reflects southern high-latitude climate²⁴, the global atmospheric methane content which reflects low-latitude climate²⁵, and the Devil's Hole isotopic record of mid-latitude climate²⁵. Most directly relevant to Greenland climate are proxy records for the deep Atlantic circulation, which show a similar duration, and a persistent interglacial maximum for approximately 10 kyr (ref. 26).

For a 'standard case' fabricated isotopic chronology before 98 kyr ago ($\Delta\delta^*$) we use the Vostok ice core deuterium history²⁴ as a template (see Methods and Fig. 1), which is scaled to the measured Greenland $\Delta\delta_E$ maximum. The $\Delta\delta^*$ yields climatic temperature changes (Methods) using the constant α_c . Using the constraints $\alpha_c = 0.15$ –0.43 per mil °C⁻¹ and $\Delta\delta_E = 3.3$ per mil, we investigated the extent to which the GIS shrank during the last interglacial. For all values $\alpha_c < 0.45$, we find that the GIS was substantially smaller than at present, and contributed more than 4 m to sea-level rise (Figs 2 and 3).

The volume reduction of the model ice sheet occurs rapidly in response to the peak warmth of the climate history. Therefore the calculated maximum sea-level contribution does not decrease

substantially if we reduce the duration of the fabricated Eemian climate by reasonable amounts (Fig. 4). In many other respects, the sea-level rise is not particularly sensitive to the imposed chronology and accumulation sensitivity (Fig. 4). This is due in part to the rapid response time (a few millennia) of the ice sheet to changes in margin position and accumulation rate, which are dominant controls on ice-sheet evolution at the timescale most relevant here. It is also due to the action of systematic buffers. Most important is the trade-off between changes of climatic temperature and those of elevation. The fixed constraint $\Delta\delta \leq 3.3$ means that if the ice-sheet summit elevation is very high (due to high accumulation rate, for example) then the climatic temperature (and hence melt rates at the margins) must also be very high. Likewise, as the summit elevation decreases in response to marginal melt, the climatic temperature must decrease too, reducing melt rates. The latter constrains the maximum sea-level contribution against increases in the duration of isotopically heavy precipitation (curves 5, 6 and 7 in Fig. 4).

Measurements of the total gas content of the GRIP core Eemian ice²⁷ provide a rough constraint on the upper limit for the sea level contribution. The highest gas contents of Eemian ice are only

slightly higher than late Holocene values, indicating that the elevation of the Eemian ice dome was not more than several hundred metres lower in elevation than at present. High frequency variability of the total gas measurement is at least $0.01 \text{ cm}^3 \text{ g}^{-1}$, which corresponds to a $\sim 350 \text{ m}$ elevation uncertainty. In our model results the corresponding constraints are $\alpha_c^i > 0.25$ and a sea-level contribution less than 5.4 m (Fig. 2).

The preceding results indicate that a sea-level contribution greater than 4 m from the GIS should be adopted as a 'most likely' scenario. Important uncertainties exist, in particular due to our lack of information about changes in the spatial pattern of accumulation rate, and also our inability to predict a priori the isotopic sensitivity. Our answer is by no means definitive, and we cannot firmly exclude a much smaller GIS contribution.

Nevertheless, the plausibility of substantial GIS melt⁶ implies that the Eemian highstand should not be invoked as evidence for major reduction of WAIS volume. Further, this Northern Hemisphere melt would provide a rather large sea-level forcing at the WAIS margins, and the Eemian history of the WAIS (unknown at present) should be viewed in this context. A more general conclusion is that the magnitude of the Eemian highstand is a poor measure of the behaviour of the ice sheets, because contributions from both hemispheres could be obscured by asynchronous changes, or by growth of the vast East Antarctic ice sheet. □

Methods

Inferring α_c^i

Temperature change at the ice core site ΔT_{obs} has components due to climatic change (ΔT_c) and elevation change (ΔZ) such that $\Delta T_{\text{obs}} = \Delta T_c + \lambda_T \Delta Z$ where λ_T is the thermal lapse rate $-7.5^\circ \text{C km}^{-1}$. The corresponding isotopic change is $\Delta\delta_{\text{obs}} = \alpha_c \Delta T_c + \alpha_z \lambda_T \Delta Z$, in which the sensitivities α_c and α_z are not necessarily equal. The α_z value depends primarily on the evolution of single air masses²⁸ and therefore, unlike α_c , may be insensitive to temporal changes of precipitation seasonality and source region temperature. We therefore make the reasonable assumption (which minimizes GIS melt in our calculations) that $\alpha_z \lambda_T$ pertaining to temporal ΔZ equals the modern spatial gradient $-6.2 \text{ per mil km}^{-1}$ (ref. 28) rather than being reduced in correlation with α_{obs} , the α value observed through time at the ice core site. Thus we calculate

$$\alpha_c = \frac{\alpha_{\text{obs}} - \alpha_z \lambda_T \Delta Z / \Delta T_{\text{obs}}}{1 - \lambda_T \Delta Z / \Delta T_{\text{obs}}} \quad (1)$$

Elevation change of the Greenland summit over the last 5 kyr is not known precisely but model-based bounds are $\Delta Z = 0$ to -100 m (refs 4, 5, 18 and 27) which gives α_c^i in the range 0.15 – 0.43 . This is lower than the 0.52 inferred for the end of the Little Ice Age²⁹, presumably because some elements of the climate system that change over millennia do not change during higher frequency climate-change events such as the Little Ice Age termination.

Mass balance parametrizations

Precipitation rates $P(x, y, t)$ are perturbed from present day ($t = t_0$) fields as a function of air temperature T , as²:

$$P(x, y, t) = P(x, y, t_0) \exp[a(T(x, y, t) - T(x, y, t_0))] \times \min \left[1.5, \frac{\nabla h(x, y, t) + 0.001}{\nabla h(x, y, t_0) + 0.001} \right] \quad (2)$$

The final factor incorporates the influence of changes in ice surface slope ∇h , of importance for orographically-driven precipitation at the margins. Present-day precipitation fields $P(x, y, t_0)$ are from refs 9 and 30. The standard case exponential factor $a = 0.0693$.

An annual mass balance is then calculated for all points and times as the difference between accumulation and melt rates. Surface melt and the fraction of precipitation to fall as snow are calculated from the annual degree-day method³. Air temperatures are assumed to be normally distributed about the monthly mean values, with a standard deviation of 5°C .

Fabricated isotopic history

Greenland Summit $\delta^{18}\text{O}$ and Vostok δD are strongly correlated for averages of several millennia or longer. In terms of the Vostok isotopic history $\Delta\delta_V$ (ref. 24) and Vostok Eemian isotopic maximum $\Delta\delta_V^E$ (Δ indicates deviations from present values) we define $\Delta\delta^* = (\Delta\delta_V / \Delta\delta_V^E) \Delta\delta_E$ (Fig. 1) for the Eemian. The preceding glacial is likewise scaled to the Vostok and Greenland minima.

Climate perturbation with isotopic lapse rate correction

The large measured value for $\Delta\delta_E$ may be partly due to reduced elevation of the ice-sheet

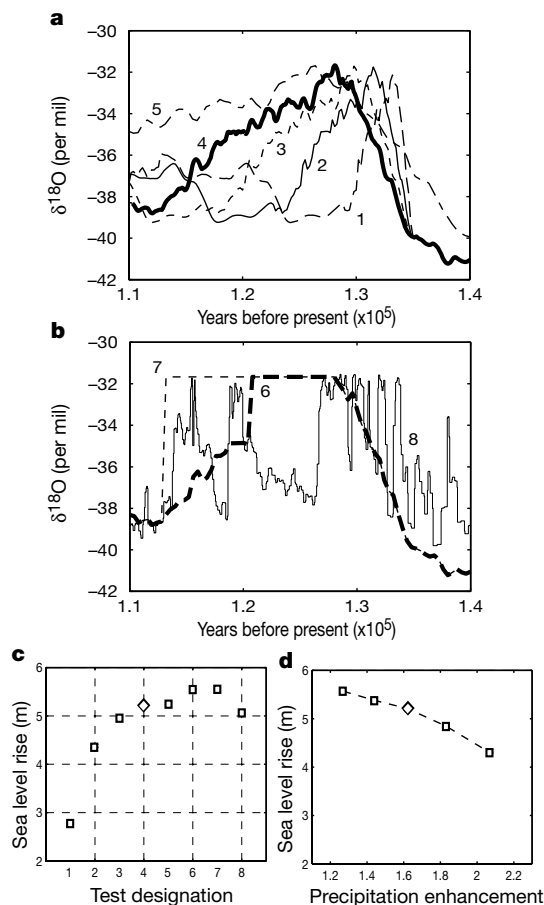


Figure 4 Sensitivity tests. Eight separate interglacial isotopic histories were used as model forcings, and calculations were performed with $\alpha_c^i = 0.33$. **a**, $\delta(t)$ for 'standard case' (Test 4) and compressed/stretched versions such that interglacial durations are $\frac{1}{4}$ (Test 1), $\frac{1}{2}$ (Test 2), $\frac{3}{4}$ (Test 3), and 2 (Test 5) times the standard case. **b**, $\delta(t)$ for a 7.5-kyr constant maximum (Test 6), a 15-kyr constant maximum (Test 7), and an unmodified GRIP forcing (Test 8). **c**, Corresponding sea-level maxima, with standard case (Test 4) indicated by a black diamond. **d**, Sensitivity to accumulation-rate parameter a , shown as sea-level maxima calculated for various values of a , which we have converted to the fractional precipitation increase due to a 7K warming. The diamond corresponds to the a value inferred from ice core studies for the glacial–interglacial transition. The average a within the Holocene interglacial climate has been markedly lower¹⁸, suggesting that the lower precipitation enhancement values are more likely.

surface. We assume that all Eemian ice beneath the modern summit originated near the high point of the Eemian ice sheet. Therefore the elevation change ΔZ is for the summit dome. Climate perturbations are then $\Delta T_c = (\Delta\delta^* - \alpha_c \lambda_T \Delta Z)/\alpha_c$. For every location on the ice-sheet surface, temperature perturbations are modulated by the thermal lapse rate and thus $T(x, y, t) - T(x, y, t_0) = \Delta T_c + \lambda_T \Delta Z(x, y, t)$. In some runs we have set $\alpha_c = \alpha_c^i$ for all ages greater than 98 kyr ago, and in others set $\alpha_c = 0.33$ for the penultimate deglaciation ($\delta^{18}\text{O} < -35$ per mil and age >120 kyr ago). The resulting difference proves to be trivial in the present context (Fig. 2).

Received 4 October 1999; accepted 4 February 2000.

1. Stirling, C. H., Esat, T. M., Lambeck, K. & McCulloch, M. T. Timing and duration of the last interglacial: evidence for a restricted interval of widespread coral reef growth. *Earth Planet. Sci. Lett.* **160**, 745–762 (1998).
2. Vezina, J., Jones, B. & Ford, D. Sea-level highstands over the last 500,000 years: evidence from the ironshore formation on Grand Cayman, British West Indies. *J. Sedim. Res.* **69**, 317–327 (1999).
3. Mercer, J. H. West Antarctic ice sheet and CO_2 greenhouse effect: a threat of disaster. *Nature* **271**, 325–325 (1978).
4. Letreguilly, A., Reeh, N. & Huybrechts, P. The Greenland ice sheet through the last glacial-interglacial cycle. *Paleogeogr. Paleoclimatol. Palaeoecol.* **90**, 385–394 (1991).
5. Ritz, C., Fabre, A. & Letreguilly, A. Sensitivity of a Greenland ice sheet model to ice flow and ablation parameters: Consequences for evolution through the last climatic cycle. *Clim. Dyn.* **13**, 11–24 (1997).
6. Koerner, R. M. Ice-core evidence for extensive melting of the Greenland Ice Sheet in the last interglacial. *Science* **244**, 964–968 (1989).
7. MacAyeal, D. R. Irregular oscillations of the West Antarctic ice sheet. *Nature* **359**, 29–32 (1992).
8. Oppenheimer, M. Global warming and the stability of the West Antarctic ice sheet. *Nature* **393**, 325–332 (1998).
9. Huybrechts, P., Letreguilly, A. & Reeh, N. The Greenland ice sheet and greenhouse warming. *Paleogeogr. Paleoclimatol. Palaeoecol. (Glob. Planet. Change)* **89**, 399–412 (1991).
10. Letreguilly, A., Huybrechts, P. & Reeh, N. Steady-state characteristics of the Greenland ice sheet under different climates. *J. Glaciol.* **37**, 149–157 (1991).
11. Cuffey, K. M. *et al.* Large Arctic temperature change at the Wisconsin-Holocene glacial transition. *Science* **270**, 455–458 (1995).
12. Dahl-Jensen, D. *et al.* Past temperatures directly from the Greenland ice sheet. *Science* **282**, 268–271 (1998).
13. Severinghaus, J. P., Sowers, T., Brook, E. J., Alley, R. B. & Bender, M. L. Timing of abrupt climate change at the end of the Younger Dryas interval from thermally fractionated gases in polar ice. *Nature* **391**, 141–146 (1998).
14. Fawcett, P. J., Agustsdottir, A. M., Alley, R. B. & Shuman, C. A. The Younger Dryas termination and North Atlantic deep water formation: Insights from climate model simulations and Greenland ice cores. *Paleoceanography* **12**, 23–38 (1997).
15. Boyle, E. A. Cool tropical temperatures shift the global $\delta^{18}\text{O}$ -T relationship: An explanation for the ice

- core borehole thermometry conflict? *Geophys. Res. Lett.* **24**, 273–276 (1997).
16. Alley, R. B., Agustsdottir, A. M. & Fawcett, P. in *Mechanisms of Millennial-Scale Global Climate Change* (eds Clark, P. U. & Webb, R. S.) (American Geophysical Union Monograph, in the press).
17. Wang, X. L., Stone, P. H. & Marotzke, J. Global thermohaline circulation. Part 1: Sensitivity to atmospheric moisture transport. *J. Clim.* **12**, 71–82 (1999).
18. Cuffey, K. M. & Clow, G. D. Temperature, accumulation and ice sheet elevation in central Greenland through the last deglacial transition. *J. Geophys. Res.* **102**, 26383–26396 (1997).
19. Grootes, P. M., Stuiver, M., White, J. W. C., Johnsen, S. & Jouzel, J. Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
20. *The Greenland Summit Ice Cores CD-ROM* (World Data Center-A for Paleoclimatology, National Geophysical Data Center, Boulder, Colorado, 1997).
21. Marshall, S. J. & Clarke, G. K. C. A continuum mixture model of ice stream thermomechanics in the Laurentide ice sheet, 1. Theory. *J. Geophys. Res.* **102**, 20599–20614 (1997).
22. Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
23. Chappellaz, J., Brook, E., Blunier, T. & Malaize, B. CH_4 and $\delta^{18}\text{O}$ of O_2 records from Antarctic and Greenland ice: A clue for stratigraphic disturbance in the bottom part of the Greenland Ice Core Project and the Greenland Ice Sheet Project 2 ice cores. *J. Geophys. Res.* **102**, 26547–26558 (1997).
24. Jouzel, J. *et al.* Vostok ice cores: extending the climatic signal over the penultimate glacial period. *Nature* **364**, 407–412 (1993).
25. Winograd, I. J. *et al.* Continuous 500,000-year climate record from vein calcite in Devils Hole, Nevada. *Science* **258**, 255–260 (1992).
26. Adkins, J. F., Boyle, E. A., Keigwin, L. & Cortijo, E. Variability of the North Atlantic thermohaline circulation during the last interglacial period. *Nature* **390**, 154–156 (1997).
27. Raynaud, D., Chappellaz, J., Ritz, C. & Martinerie, P. Air content along the Greenland Ice Core Project core: A record of surface climatic parameters and elevation in central Greenland. *J. Geophys. Res.* **102**, 26607–26613 (1997).
28. Johnsen, S. J., Dansgaard, W. & White, J. W. C. The origin of Arctic precipitation under present and glacial conditions. *Tellus* **41**, 452–469 (1989).
29. Cuffey, K. M., Alley, R. B., Grootes, P. M., Bolzan, J. M. & Anandakrishnan, S. Calibration of the $\delta^{18}\text{O}$ isotopic paleothermometer for central Greenland, using borehole temperatures. *J. Glaciol.* **40**, 341–349 (1994).
30. Ohmura, A. & Reeh, N. New precipitation and accumulation maps for Greenland. *J. Glaciol.* **37**, 140–148 (1991).

Acknowledgements

We thank C. Ritz, G. Clow, National Science and Engineering Research Council Canada, GISP2 and GRIP project members, and especially the Northwest Glaciological Society.

Correspondence and requests for materials should be addressed to K.C. (e-mail: kcuffey@socrates.berkeley.edu).

KNOW YOUR COPY RIGHTS RESPECT OURS

The publication you are reading is protected by copyright law.
Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

Rapid and early export of *Phaeocystis antarctica* blooms in the Ross Sea, Antarctica

G. R. DiTullio*, J. M. Grebmeier†, K. R. Arrigo‡, M. P. Lizotte§, D. H. Robinson||, A. Leventer¶, J. P. Barry#, M. L. VanWoert* & R. B. Dunbar**

* University of Charleston, Grice Marine Laboratory, 205 Fort Johnson, Charleston, South Carolina 29412, USA

† University of Tennessee, Knoxville, Tennessee 37996, USA

‡ Department of Geophysics, Stanford University, Stanford, California 94305-2215, USA

§ Department of Biology and Microbiology, University of Wisconsin Oshkosh, Oshkosh, Wisconsin 54901-8460, USA

|| Romberg Tiburon Center, San Francisco State University, P.O. Box 855, Tiburon, California 94920-0855, USA

¶ Department of Geology, Colgate University, Hamilton, New York 13346, USA

Monterey Bay Aquarium Research Institute, Moss Landing, California 95039-0628, USA

* NOAA/NESDIS/Office of Research and Applications, E/RA3, NOAA Science Center Room 711, 5200 Auth Road, Camp Springs, Maryland 20746, and The National Ice Center, Federal Office Building 4, Room 2301, 4251 Suitland Road, Washington DC 20395, USA

** Geophysics and Environmental Sciences, Stanford University, Stanford, California 94305-2115, USA

The Southern Ocean is very important for the potential sequestration of carbon dioxide in the oceans¹ and is expected to be vulnerable to changes in carbon export forced by anthropogenic climate warming². Annual phytoplankton blooms in seasonal ice zones are highly productive and are thought to contribute sig-

nificantly to pCO₂ drawdown in the Southern Ocean. Diatoms are assumed to be the most important phytoplankton class with respect to export production in the Southern Ocean; however, the colonial prymnesiophyte *Phaeocystis antarctica* regularly forms huge blooms in seasonal ice zones and coastal Antarctic waters³. There is little evidence regarding the fate of carbon produced by *P. antarctica* in the Southern Ocean, although remineralization in the upper water column has been proposed to be the main pathway in polar waters^{4,5}. Here we present evidence for early and rapid carbon export from *P. antarctica* blooms to deep water and sediments in the Ross Sea. Carbon sequestration from *P. antarctica* blooms may influence the carbon cycle in the Southern Ocean, especially if projected climatic changes lead to an alteration in the structure of the phytoplankton community^{6,7}.

As part of the Research on Ocean–Atmosphere Variability and Ecosystem Response in the Ross Sea (ROAVERRS) programme, we found extensive blooms of *Phaeocystis antarctica*, a colonial prymnesiophyte. *P. antarctica* allocates ~50% of its carbon to the extracellular mucus that makes up its colonial matrix⁸. Nutrients drawn down in regions dominated by *P. antarctica* have ratios of carbon to phosphorus that are significantly greater than the Redfield (C/P = 106) ratio, indicating the greater potential of *P. antarctica* to export carbon relative to diatoms^{6,7}. It has been assumed that senescing *Phaeocystis* blooms are remineralized within the upper water column (depth <100 m) and hence do not influence carbon sequestration⁵. Here we suggest, however, that in the Ross Sea *P. antarctica* blooms are exported rapidly during the early spring bloom.

Before ROAVERRS NBP-96-6 sampling (18 December 1996 to 8 January 1997), satellite images of ocean colour⁷ showed a large phytoplankton bloom along the Ross Ice Shelf (RIS) as early as 8 November 1996. However, images collected only five days later showed that levels of chlorophyll associated with this bloom had already declined. We measured (24–26 December 1996) the integral

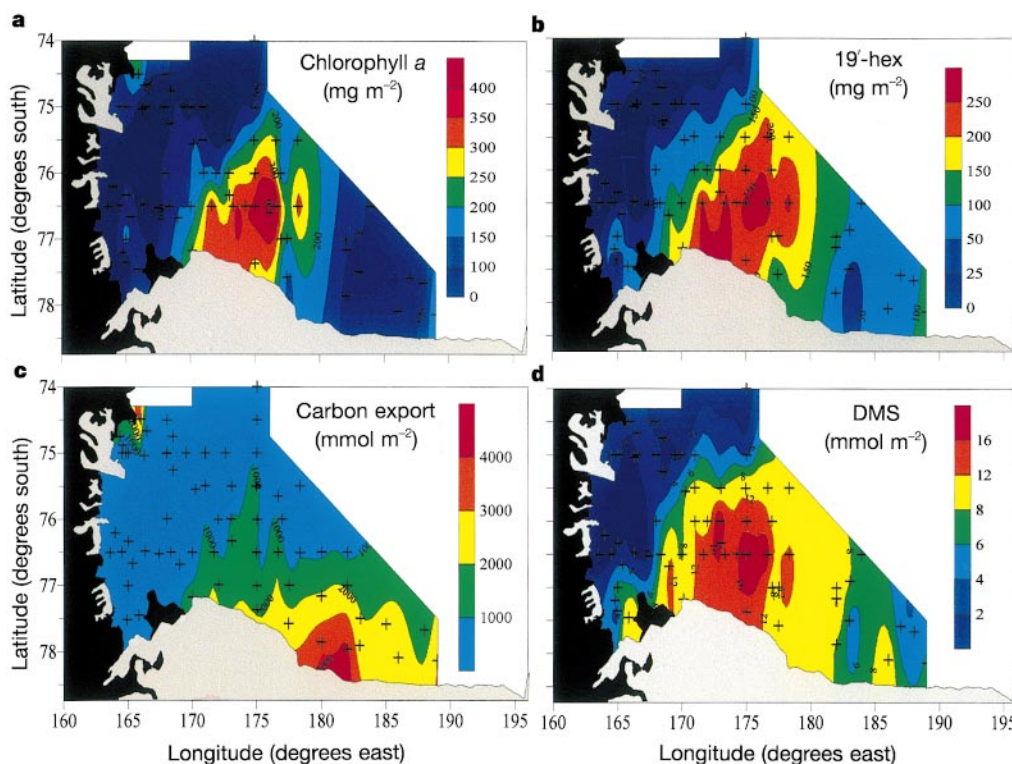


Figure 1 Integral values (depth 0–150 m) of chlorophyll *a* (a), 19'-hexanoyloxyfucoxanthin (19'-hex) (b), carbon export (c) and dimethylsulphide (DMS) (d) measured during the ROAVERRS-96-6 cruise (18 December 1996 to 8 January 1997). Carbon export was

estimated from the carbon deficit (Δ TDIC – POC) in the water column assuming a winter¹¹ TDIC value of 2,287 mmol m⁻³.

chlorophyll values to be 25% of those measured in the central bloom area (>400 mg chlorophyll per m^2 ; Fig. 1). Fluorescence profiles from this region showed deep (400–550 m) phytoplankton peaks at several stations along the RIS. These fluorescence peaks were associated with dimethylsulphide (DMS) concentrations up to 300-fold higher (4–15 nM; Fig. 2) than those previously measured in the deep waters (depth >400 m) of the Ross Sea⁹. DMS is an effective chemical indicator of the recent presence of *P. antarctica* in the Ross Sea⁹. High *P. antarctica* cell concentrations (about 5×10^6 cells per litre) were observed in this deep water (Fig. 2), demonstrating that it was not simply the colonial matrix that sedimented out of the upper water column in the Ross Sea, as was observed for *Phaeocystis pouchetii* blooms in the Balsford¹⁰.

The photochemical quantum efficiency of photosystem II (variable/maximal fluorescence (F_v/F_m)) for *P. antarctica* cells obtained from 450–500 m depth ranged from 0.20 to 0.25 (Fig. 2). Although these values indicate some physiological stress, the similarity to surface values (about 0.3–0.4) indicates that the photosynthetic apparatus of deep *P. antarctica* cells was not significantly degraded relative to cells at the surface. The active physiological state of this deep *P. antarctica* population implies that they were rapidly exported from nearby surface waters. High sinking speeds (much greater than 200 m day^{-1}) have been measured for *P. antarctica* aggregates during bloom conditions in the Ross Sea⁴, meaning that cells could theoretically reach deep water in only 1 to 2 days. This rapid export probably accounts for the physiological viability of the deep water cells. Relatively low current velocities (<10 cm s^{-1}) measured at nearby instrument moorings and a sharp pycnocline imply that water mass advection and convection were relatively unimportant with respect to vertical flux, and that sinking was the main cause of the export.

In December 1996, surface NO_3^- concentrations measured near the RIS were reduced from their winter¹¹ values of ~ 30 μM to 11 μM . It is likely that seasonal iron limitation terminated this bloom¹². The silicate concentrations in these waters were unchanged from their winter¹¹ values (about 76 μM), indicating that most of the NO_3^- drawdown was due to *P. antarctica* rather than to diatoms. Only $\sim 20\%$ of the CO_2 drawdown, as estimated from the change in

total dissolved inorganic carbon ($\Delta TDIC$) in the upper 150 m of the water column, could be accounted for by suspended particulate organic carbon (POC) (Fig. 1). On the basis of previous measurements, it is unlikely that either dissolved organic carbon (DOC) production¹³ or zooplankton grazing^{14,15} contributed significantly to the POC deficit in the water column along the RIS. Instead, ROAVERRS data indicate that this deficit may have been due to the rapid vertical export of carbon between 8 November and 24 December 1996.

Perhaps the best evidence for the rapid export of carbon by *P. antarctica* comes from the concentrations of particulate dimethylsulphoniopropionate (DMSP_p; an intracellular precursor of DMS) measured in surface sediments. Hydrolysis of DMSP produces DMS and acrylic acid in equimolar ratios¹⁶ through grazing¹⁷ and DMSP-lyase activity¹⁸. DMSP_p concentrations in the sediments of the Ross Sea can be used as a diagnostic indicator of *P. antarctica* export from surface waters. DMSP_p concentrations along the RIS were about 15 μg DMSP per gram dry sediment weight, equivalent to ~ 300 mg DMSP per m^2 or 105 mg carbon per m^2 as DMSP. These concentrations represent about 1.75 g carbon per m^2 of *P. antarctica* carbon (assuming that 6% of carbon in *P. antarctica* cells are present as DMSP (ref. 19)). Because some fraction of the *P. antarctica* export flux is remineralized during export^{4,5} or at the sediment–water interface, the estimates of carbon-flux to the ocean floor on the basis of the DMSP_p content of the sediment represent a lower boundary to the true carbon export value associated with *P. antarctica* blooms. However, sediment focusing in this region could complicate the interpretation of the areal DMSP_p fluxes²⁰.

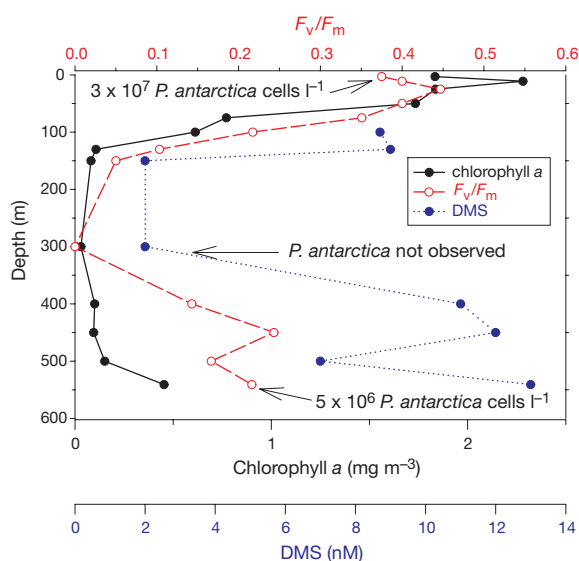


Figure 2 Vertical profiles of *P. antarctica* cell numbers, chlorophyll *a* and DMS concentrations, and F_v/F_m near the Ross Ice Shelf ($76.1^\circ S$, $174^\circ W$) on 24 December 1996. DMS concentrations in the upper 100 m of the water column were not included as these values were more than an order of magnitude higher than those for samples taken below a depth of 100 m. Note that the DMS minimum layer corresponds to the F_v/F_m and chlorophyll *a* minimum layers.

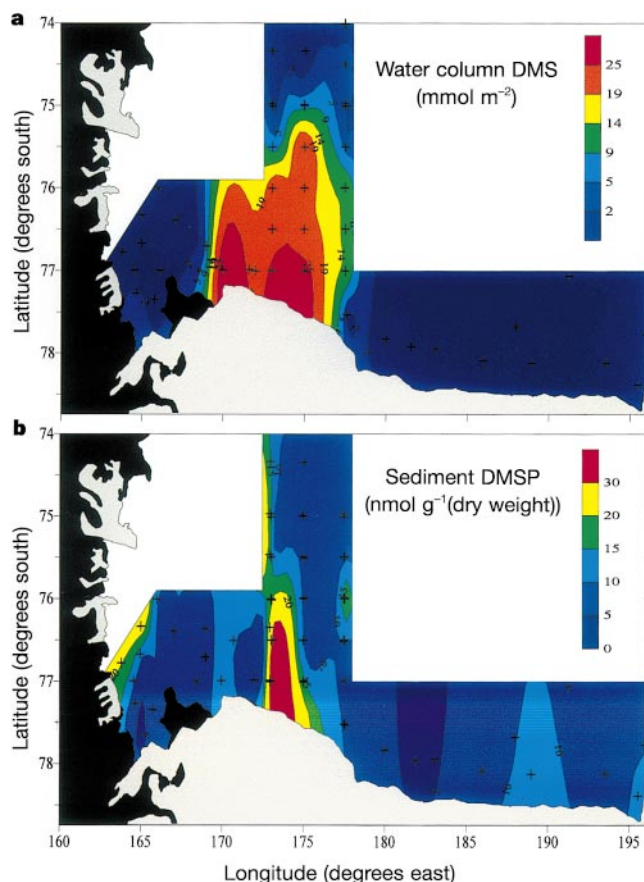


Figure 3 Spatial distribution of DMS concentrations in the integral water column (depth = 0–150 m) (a) and DMSP concentrations in the surface sediment (depth = 0–2 cm) (b) measured during the early spring bloom period (November 1998) in the Ross Sea (ROAVERRS-98-7).

In November 1998, we observed evidence for the export of *P. antarctica* from surface waters in early blooms. DMSP_p concentrations in surface sediments were more than twice as high in November 1998 as in December 1996 (Fig. 3), implying a carbon export due to *P. antarctica* of ~4 g carbon per m². Fluorescence profiles and elevated DMS concentrations at various depths (300–650 m) from several stations in the central Ross Sea revealed the presence of deep *P. antarctica* populations. More importantly, a strong spatial correlation was observed between areal sediment DMSP and the developing *P. antarctica* bloom (Fig. 3). Although some lateral advection undoubtedly occurs, vertical export must be significantly faster than lateral processes to explain this strong spatial relationship (Fig. 3). Deep water fluorescence signals with elevated DMS concentrations were observed during four consecutive field seasons at various locations in the Ross Sea, indicating the common occurrence of *P. antarctica* flux events.

High carbon flux during early bloom conditions implies that senescence was not a prerequisite for *P. antarctica* export to deep water and sediments. Hence, carbon export from *P. antarctica* blooms in the Ross Sea was not dependent on the demise of the bloom itself. Moreover, because DOC concentrations and bacterial production in the spring are low in the Ross Sea¹³, and because of the rapid sinking speeds of aggregates⁴, a significant fraction of the carbon export associated with *P. antarctica* must escape remineralization in the upper water column (depth <100 m) and be exported to deep waters. On the basis of the carbon export that was recorded in November 1998, the large carbon deficit that was observed along the RIS during December 1996 probably did not result from a single episodic event due to bloom senescence. Rather, the numerous fluorescence peaks that were observed at various depths indicate that carbon export due to *P. antarctica* blooms results from multiple episodic export events. Also, rates of oxygen uptake by the sediment (which indicate carbon flux to the ocean floor) were at their highest in 1998 in the same region as the high-sediment DMSP. Interannual variability in the locations of regions where levels of oxygen uptake by the sediment are high is also consistent with multiple episodic export events.

Sediment traps were not installed adjacent to the central or eastern RIS from 1996 to 1998 as part of the ROAVERRS and US JGOFS (United States Joint Global Ocean Flux Study) field programmes. However, sediment trap fluxes were measured between January 1995 and January 1996 at 423 m at a site adjacent to the RIS in the central Ross Sea (78° S, 177° W) (ref. 21). The highest daily average for POC flux during the year occurred between 8 December and 23 December 1995, at a time when the trapped particles had relatively high organic carbon to silicon ratios. These findings are consistent with our own observations of *Phaeocystis* export.

Our data may have a direct bearing on the glacial iron hypothesis²² because of the potential link between *Phaeocystis* and the biological pump. A recent study of the seasonal ice zones during the Last Glacial Maximum (LGM) concluded that biological production in the Southern Ocean contributed substantially to the lowering of atmospheric pCO₂ (ref. 23). In addition, radionuclide proxies have shown that POC export in the subantarctic zone was significantly higher during the LGM than during the Holocene epoch²⁴. Because silicate utilization by diatoms was diminished during the LGM relative to the present interglacial²⁵, it is possible that a non-siliceous phytoplankton species may have been important at this time. Considering the high productivity²⁶ and carbon-export potential of *P. antarctica* blooms, this species may have been important during the LGM in sequestering atmospheric pCO₂ in the Southern Ocean. On the basis of the high levels of methanesulphonic acid (a product of the atmospheric breakdown of DMS) during the LGM²⁷ and its positive correlation with proxies of iron in ice cores²⁸, it seems that *P. antarctica* may represent an important link in the carbon, iron and sulphur cycles. Furthermore, our results indicate that the episodic sinking of colonial *P. antarctica* blooms

may not require bloom senescence conditions, a fact that may facilitate the continuous drawdown of pCO₂ during the austral summer. Thus, early bloom conditions in the Ross Sea may be important with respect to carbon export from surface waters, particularly as *P. antarctica* blooms remove about twice as much CO₂ per mole of PO₄ relative to diatoms⁷. The next generation of ocean models must incorporate the possible effects of algal species composition on carbon export in the Southern Ocean, especially if rising levels of CO₂ lead to shifts in the composition of the phytoplankton community^{6,7}. □

Methods

Water samples were collected using 10-litre Bullister Bottles attached to a conductivity temperature depth (CTD) rosette. Pigment concentrations were measured using high-performance liquid chromatography (HPLC) pigment analyses⁹. Concentrations of 19'-hexanoyloxyfucoxanthin are a proxy for the presence of *P. antarctica* in the Ross Sea⁹. The *F_v/F_m* measurements were made using a pulse-amplitude modulated fluorometer. Cells were adapted to the dark for 30 min before measurement. *P. antarctica* cell counts were determined by microscopy and the bright autofluorescence of cells at depth implied rapid vertical export as opposed to lateral advection. Surface sediments were collected with both box and Haps corers and sampled in duplicate at various sites on the core. Most sampling sites were reoccupied and the average sediment DMSP value was used. DMS concentrations were measured using cryogenic purge and trap gas chromatography⁹. Aliquots of sediment were subjected to base hydrolysis (6N NaOH) for 24 h in the dark to convert DMSP to DMS¹⁸ and were analysed at sea. Sediments were dried and weighed in centrifuge tubes for normalization. For our estimate of the carbon deficit (Fig. 1), we assumed that DOC concentrations were constant during the early spring bloom period¹³.

Received 23 September 1999; accepted 28 January 2000.

- Sarmiento, J. L. & Le Que, C. Ocean carbon dioxide uptake in a model of century-scale global warming. *Science* **274**, 1346–1350 (1996).
- Sarmiento, J. L., Hughes, T. M. C., Stouffer, R. J. & Manabe, S. Simulated response of the ocean carbon cycle to anthropogenic climate warming. *Nature* **393**, 245–249 (1998).
- El-Sayed, S. Z., Biggs, D. C. & Holm-Hansen, O. Phytoplankton standing crop, primary productivity and near-surface nitrogenous nutrient fields in the Ross Sea, Antarctica. *Deep-Sea Res.* **30**, 871–886 (1993).
- Asper, V. & Smith, W. O. Jr Particle fluxes during austral spring and summer in the southern Ross Sea, Antarctica. *J. Geophys. Res.* **104**, 5345–5359 (1999).
- Wassman, P., Vernet, M., Mitchell, B. G. & Rey, F. Mass sedimentation of *Phaeocystis pouchetii* in the Barents Sea. *Mar. Ecol. Prog. Ser.* **66**, 183–195 (1990).
- Arrigo, K. R. et al. Phytoplankton community structure and the drawdown of nutrients and CO₂ in the Southern Ocean. *Science* **283**, 365–367.
- Arrigo, K. R. et al. Phytoplankton taxonomic variability in nutrient utilization and primary production in the Ross Sea. *J. Geophys. Res.* (in the press).
- Matrai, P. A. et al. Light dependence of carbon and sulfur production by polar clones of the genus *Phaeocystis*. *Mar. Biol.* **124**, 157–167 (1995).
- DiTullio, G. R. & Smith, W. O. Jr Relationship between dimethylsulfide and phytoplankton pigment concentrations in the Ross Sea, Antarctica. *Deep-Sea Res.* **42**, 873–892 (1995).
- Passow, U. & Wassman, P. On the trophic fate of *Phaeocystis pouchetii* (Hariot): I.V. The formation of snow by *P. pouchetii*. *Mar. Ecol. Prog. Ser.* **104**, 153–161 (1994).
- Sweeney, C. et al. The carbon budget for the Ross Sea: A model for the Southern Ocean. *Deep-Sea Res.* (in the press).
- Sedwick, P. N., DiTullio, G. R. & Mackey, D. J. Iron and manganese in the Ross Sea, Antarctica: Seasonal iron limitation in Antarctic shelf waters. *J. Geophys. Res.* (in the press).
- Carlson, C. A., Ducklow, H. W., Hansell, D. A. & Smith, W. O. Jr Organic carbon partitioning during spring phytoplankton blooms in the Ross Sea polynya and the Sargasso Sea. *Limnol. Oceanogr.* **43**, 375–386 (1998).
- Hansen, E. C. & van Boekel, W. H. M. Grazing pressure of the calanoid copepod *Temora longicornis* on a *Phaeocystis*-dominated spring bloom in a Dutch tidal inlet. *Mar. Ecol. Prog. Ser.* **78**, 123–129 (1991).
- Verity, P. G. & Smayda, T. J. Nutritional value of *Phaeocystis pouchetii* (Prymnesiophyceae) and other phytoplankton for *Arctia* spp. (Copepoda): Ingestion, egg production, and growth of nauplii. *Mar. Biol. Berlin* **100**, 161–171 (1989).
- Cantoni, G. L. & Anderson, D. G. Enzymatic cleavage of dimethylpropiothetin by *Polysiphonia lanosa*. *J. Biol. Chem.* **222**, 171–177 (1956).
- Dacey, J. W. H. & Wakeham, S. G. Oceanic DMS: production during zooplankton grazing on phytoplankton. *Science* **233**, 1314–1316 (1986).
- Stefels, J., Gieskes, W. W. C. & Dijkhuizen, L. in *Biological and Environmental Chemistry of DMSP and Related Sulfonium Compounds* (eds Kiene, R. P., Visscher, P. T., Keller, M. D. & Kirst, G. O.) 305–316 (Plenum, New York, 1996).
- Stefels, J. & van Leeuwe, M. A. Effects of iron and light stress on the biochemical composition of Antarctic *Phaeocystis* sp. (Prymnesiophyceae). I. Intracellular DMSP concentrations. *J. Phycol.* **34**, 486–495 (1998).
- Dunbar, R. B., Anderson, J. B., Domack, E. W. & Jacobs, S. S. in *Oceanology of the Antarctic Shelf* (ed. Jacobs, S. S.) Antarctic Research Series Vol. 43, 291–312 (American Geophysical Union, Washington DC, 1985).
- Accornero, A., Bergamasco, A., Monaco, A. & Tucci, S. in *Oceanography of the Ross Sea, Antarctica* (eds Spezie, G. & Mauzella, G. M. R.) 177–196 (Springer, Milan, 1999).
- Martin, J. H. Glacial-interglacial CO₂ change: The iron hypothesis. *Paleoceanography* **5**, 1–13 (1990).
- Moore, J. K., Abbott, M. R., Rich, J. G. & Nelson, D. H. The Southern Ocean at the last glacial maximum: A strong sink for atmospheric carbon dioxide. *Glob. Biogeochem. Cycles* (in the press).
- Kumar, N. et al. Increased biological productivity and export production in the glacial Southern

- Ocean. *Nature* **378**, 675–680 (1995).
25. De La Rocha, C. L., Brzezinski, M. A., DeNiro, M. J. & Shemesh, A. Silicon-isotope composition of diatoms as an indicator of past oceanic change. *Nature* **395**, 680–683 (1998).
26. Smith, W. O. Jr & Gordon, L. I. Hyperproductivity of the Ross Sea (Antarctica) polynya during austral spring. *Geophys. Res. Lett.* **24**, 233–236 (1997).
27. LeGrand, M. et al. Ice-core record of oceanic emissions of dimethylsulfide during the last climate cycle. *Nature* **350**, 144–146 (1991).
28. Martin, J. H. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 123–128 (Plenum, New York, 1992).

Acknowledgements

We thank D. Jones, M. Geesey and J. Maucher for technical assistance with DMS/P and HPLC pigment analyses and data reduction. We also thank J. Borkowski and the crew of the RVIB *Nathaniel B. Palmer*.

Correspondence and requests for materials should be addressed to G.R.D. (e-mail: ditulioj@cofc.edu).

Developmental cheating in the social bacterium *Myxococcus xanthus*

Gregory J. Velicer^{*†}, Lee Kroos^{*} & Richard E. Lenski[†]

^{*} Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824, USA

[†] Center for Microbial Ecology, Michigan State University, East Lansing, Michigan 48824, USA

Cheating is a potential problem in any social system that depends on cooperation and in which actions that benefit a group are costly to individuals that perform them^{1–5}. Genetic mutants that fail to perform a group-beneficial function but that reap the benefits of belonging to the group should have a within-group selective advantage, provided that the mutants are not too common. Here we show that social cheating exists even among prokaryotes. The bacterium *Myxococcus xanthus* exhibits several social behaviours, including aggregation of cells into spore-producing fruiting bodies during starvation. We examined a number of *M. xanthus* genotypes that were defective for fruiting-body development, including several lines that evolved for 1,000 generations under asocial conditions⁶ and others carrying defined mutations in

developmental pathways^{7–10}, to determine whether they behaved as cheaters when mixed with their developmentally proficient progenitor. Clones from several evolved lines and two defined mutants exhibited cheating during development, being over-represented among resulting spores relative to their initial frequency in the mixture. The ease of finding anti-social behaviours suggests that cheaters may be common in natural populations of *M. xanthus*.

The myxobacteria are soil-dwelling prokaryotes that exhibit a social life cycle similar to that of the eukaryotic slime mould *Dictyostelium discoideum*^{11,12}. Like the slime moulds, myxobacteria undergo multicellular development in response to starvation that yields fruiting bodies of remarkable morphological variety¹³. This process involves multiple intercellular signals specific to distinct stages of development, and it results in a minority of the original population becoming stress-resistant spores that germinate under favourable conditions¹⁴. Myxobacteria also exhibit social motility and social predation, swarming as a ‘wolf pack’ toward prey, which they kill and degrade by the secretion of extracellular compounds¹⁵.

A developmentally defective mutant of *M. xanthus*, when mixed with a developmentally proficient wild type, has five possible fates during development. First, the defective strain may sporulate with the same efficiency as it does in pure culture. This outcome corresponds to null hypothesis H₁ here. Second, a partially defective genotype’s sporulation in the presence of wild type may be inhibited even below its efficiency in pure culture. The third and fourth potential fates are partial and complete rescue (relative to wild type), respectively, of the defective genotype by extracellular complementation in the presence of wild type. Complete rescue to the wild-type sporulation efficiency corresponds to null hypothesis H₂ here. Fifth, a developmentally defective genotype may produce more spores in the presence of wild type than would a neutrally marked wild type introduced at the same initial frequency. This last outcome constitutes evolutionary cheating, because the defective mutant obtains disproportionate reproductive success.

We first compared the developmental performance, in pure culture, of six experimentally evolved clones with that of their wild-type ancestor DK1622 (ref. 16). In pure culture, all six clones, to varying degrees, showed defects in spore production relative to DK1622 (Fig. 1). Thus, these evolved clones were all defective for this group function. We then measured spore production of these evolved defective clones when they were each mixed with their ancestor at an initial frequency of 0.01. The performance of each minority genotype was then contrasted with the two distinct hypothetical outcomes, H₁ and H₂ (Table 1). For five out of the six

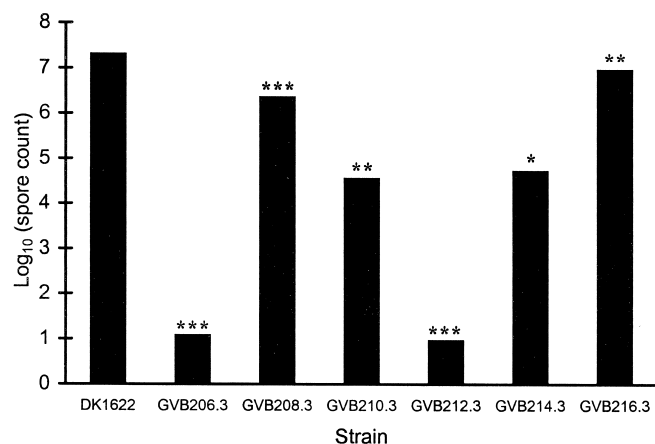


Figure 1 Spore production for six independently evolved clones⁶ of *M. xanthus* and their common ancestor (DK1622). Asterisks indicate significant defects in an evolved clone’s spore production relative to that of DK1622, calculated using Welch’s approximate *t*-test which does not assume homogeneity of variance. * *P* < 0.05; ** *P* < 0.01; *** *P* < 0.001, all one-tailed.

Table 1 Observed spore production of evolved clones when mixed at 1% with the wild-type ancestor compared with production expected under two hypotheses

Evolved clone	Observed	H ₁ expected	Observed — H ₁ expected	Observed — H ₂ expected
GVB206.3	–0.31	–7.30	+6.99***	+1.69***
GVB208.3	–0.98	–2.96	+1.98***	+1.02***
GVB210.3	–3.32	–4.75	+1.43**	–1.32**
GVB212.3	–3.27	–7.30	+4.03**	–1.27*
GVB214.3	–0.93	–4.59	+3.66***	+1.07**
GVB216.3	–3.49	–2.35	–1.14***	–1.49***

All values are log₁₀-transformed; an observed value of –3 thus indicates that an evolved clone produced 0.1% of the pure-culture wild-type spore count. According to H₁, a clone produces spores with the same efficiency when mixed with the ancestor as it does when alone. For example, a pure culture of GVB208.3 produces ~10% as many spores as does a pure culture of the wild type (Fig. 1). If GVB208.3 makes spores with the same 10% relative efficiency when mixed with its ancestor at a frequency of 1%, then it will produce 0.1% of the total spores. Expected values under H₁ vary among the evolved clones; a log₁₀-transformed value of –7.30 represents the limit of detection. A positive difference between the observed spore production of an evolved clone and its expectation under H₁ indicates at least partial complementation of the evolved clone by the wild-type. According to H₂, an evolved clone behaves as would a neutrally marked variant of the wild-type; that is, it produces 1% of the total spores when mixed at 1% with the wild-type ancestor. The expected value under H₂ is thus –2 (log₁₀ 0.01) for all clones. A positive difference between an evolved clone’s observed spore production and the uniform expectation under H₂ demonstrates developmental cheating by the evolved clone. Asterisks denote that contrasts are significantly different from zero based on *t*-tests: * *P* < 0.05; ** *P* < 0.01; *** *P* < 0.001, all two-tailed.

defective clones, spore production was higher than predicted under H_1 , thus demonstrating at least partial complementation of their developmental defects. All five of these clones were rescued more than 10-fold above the expectation under H_1 , and in three cases the effect was greater than 1,000-fold. Sporulation of the sixth clone (GVB216.3) was hindered by the wild-type majority.

Out of the five clones exceeding the sporulation expected under H_1 , two showed partial complementation, because they performed worse than expected under H_2 ; however, three clones exhibited cheating, as they performed significantly better than expected even under H_2 (Table 1). Two of these cheaters (GVB208.3 and GVB214.3) showed moderate sporulation efficiencies as pure genotypes, but they had an advantage over wild type when they were rare. Clone GVB206.3 showed the most marked cheating: it was almost completely defective at sporulation in pure culture, but as a 1% minority it produced ~50-fold more spores than would a neutrally marked wild type.

To investigate more closely the frequency-dependent behaviour of these interactions, we measured the sporulation efficiencies of two evolved cheater clones (GVB206.3 and GVB208.3) in mixes with their wild-type progenitor at nine different initial frequencies (Fig. 2). Both clones cheated by performing better than expected for a neutrally marked wild type (H_2) over a wide range of initial frequencies (Fig. 2a, b). Notably, as the initial frequency of each cheater genotype increased, the total spore production of the mixed culture fell below that of the pure wild type; this result indicates that cheaters indeed harm their group's performance. Both cheaters sporulated much more efficiently than wild type at low initial frequencies, but their efficiencies dropped below the wild-type efficiency at high initial ratios (Fig. 2c, d). The relative sporulation efficiency dropped below 1 at a lower initial frequency for GVB206.3 than for GVB208.3.

The evolved defective clones may have several mutations that distinguish them from their wild-type ancestor⁶. It is therefore unclear whether the cheating phenotype can arise by a single mutation or whether it requires multiple mutations. To test whether a single mutation can produce cheating behaviour, we examined three genotypes that differ from their wild-type progenitor by a defined mutation. The production of *M. xanthus* fruiting bodies

involves several extracellular signals that are expressed at specific stages of development^{17,18}, and the defined mutants that we examined are defective in production of signal molecules. Sporulation of these mutants is defective (at least partially) in pure culture, but can be rescued (at least partially) by extracellular complementation when mutants are mixed at a 1:1 ratio with wild type. Our re-analysis of published data^{9,10,19} from such mixes suggested that some mutants may even exhibit cheating behaviour. Therefore, we mixed a defined mutant that is defective in the production of C-signal with its wild-type progenitor at 1:1 and 1:99 ratios, and we carried out corresponding experiments with two different mutants defective in A-signal production.

One of the A-signal mutants (MS2021)¹⁰, which has a mutation in the *asgE* gene, showed partial complementation but no cheating (data not shown); that is, it fell between the expectations under H_1 and H_2 . The C-signal mutant (LS523)⁸ and the other A-signal mutant (DK4312)⁷, which has a mutation in the *asgB* gene, both showed strong cheating behaviour when mixed with their respective wild-type progenitors. DK4312 was severely defective for sporulation in pure culture, but performed much better than expected for a neutrally marked wild type (H_2) at an initial frequency of 0.01 (Fig. 3a). LS523, which also sporulated poorly in pure culture, showed a similar degree of cheating at an initial frequency of 0.01 (Fig. 3b). In 1:1 mixes, both mutants dominated the wild type and the total spore counts fell below wild-type levels, with LS523 having a greater negative effect on total spore yield than DK4312 (data not shown). These defined mutants failed to contribute normal amounts of a particular developmental signal to the group, yet they performed better than wild type when rare. Therefore, even single mutations can produce cheating behaviour in *M. xanthus*.

Our experiments with both the evolved clones and defined mutants indicate that developmental cheaters are readily obtained in *M. xanthus*. Cheater genotypes presumably also appear in nature. (Although cheaters have not been studied in natural populations of *M. xanthus*, developmental cheating has been reported in a natural isolate of the eukaryotic slime mould *Dictyostelium mucoroides*²⁰.) Once present, cheaters have the potential to invade social groups and persist indefinitely as a parasitic subpopulation. Our results therefore suggest that cheaters of myxobacteria are common in

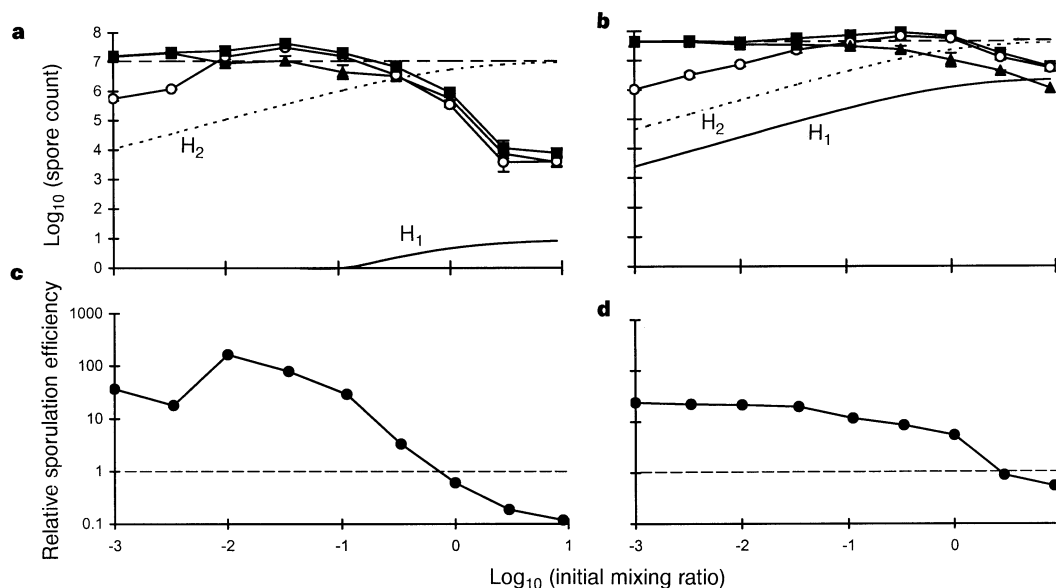


Figure 2 Spore production of two evolved clones when mixed with their wild-type progenitor at nine initial ratios, and the corresponding relative sporulation efficiencies. **a**, Mixtures of GVB206.3 with DK1622. **b**, Mixtures of GVB208.3 with DK1622. Squares, triangles and circles indicate total, DK1622 and evolved clones, respectively. The expected production of the evolved clones under H_1 (solid lines), the expected production

of evolved clones under H_2 (dotted lines) and the spore production of DK1622 in independent pure cultures (dashed lines) are shown. Error bars indicate 95% confidence intervals. **c, d**, Sporulation efficiencies of GVB206.3 (**c**) and GVB208.3 (**d**), relative to that of DK1622 for these same initial mixing ratios. Dashed lines indicate a relative efficiency of 1.

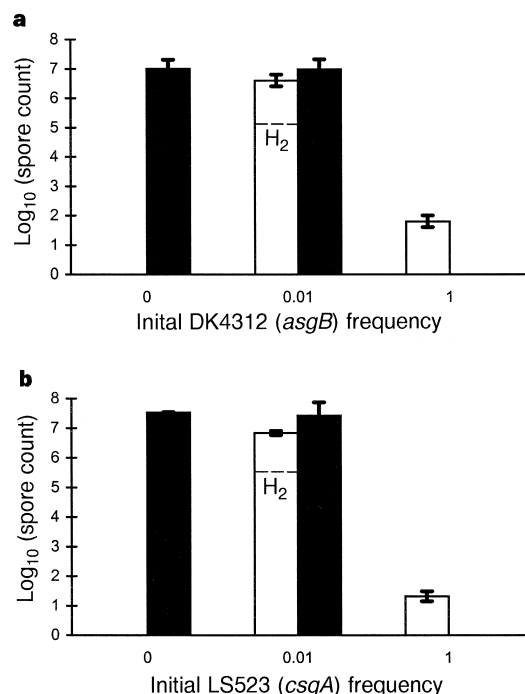


Figure 3 Spore production of mutants DK4312 (*asgB*) (**a**) and LS523 (*csgA*) (**b**), and their respective wild-type progenitors in pure and mixed culture. Dashed lines show the expected production of each mutant as a 1% minority under hypothesis H_2 , a value above which indicates developmental cheating. Error bars indicate 95% confidence intervals.

nature, unless these bacteria have evolved strategies that oppose cheating. Such strategies may include reproductive cycles that are unfavourable to cheaters and policing functions that repress competition^{21–27}.

In principle, the myxobacterial reproductive cycle could involve the founding of new colonies by either single or multiple spores. With single-spore colonization, the outcome of competition between cheaters and cooperators will depend on the performance of genetically homogeneous groups (assuming no migration or appearance of new mutants), a situation in which the developmentally defective cheaters are inferior. Alternatively, if colonies are founded by clumps of many spores, competition will occur within fruiting bodies, a situation in which cheaters are superior when rare and will therefore persist as a parasitic subpopulation. Two features of the myxobacteria suggest they do not exclude cheaters by single-spore dispersal. First, *M. xanthus* spores adhere tightly to one another and require vigorous sonication to separate them in the laboratory. Thus, new colonies in nature are probably founded by multiple clumped spores. Second, myxobacteria are highly motile, promoting migration that opposes the homogenizing effect of single-spore colonization. Barring some unknown policing mechanism that represses cheater genotypes, we therefore predict that cheaters are common in natural populations of *M. xanthus*. □

Methods

Strains

Two wild-type clones of *M. xanthus* were used in this study, along with mutant genotypes derived from each of them. Wild-type strain DK1622 (ref. 16) is sensitive to the antibiotics used to distinguish competing genotypes. DK1622 is ancestral to the evolved genotypes GVB206.3, GVB208.3, GVB210.3, GVB212.3, GVB214.3 and GVB216.3, all of which are genetically marked by resistance to rifampicin⁶. These six clones were isolated from six populations that had evolved independently from DK1622 for 1,000 generations in a nutrient-rich, liquid habitat⁶. The C-signal defective mutant LS523 (*csgA* mutation) also derives from DK1622 through genetic manipulation and is resistant to oxytetracycline⁸. DK101 is another developmentally wild-type strain that carries a *pilQ* mutation that hinders S motility^{28,29}; it is sensitive to kanamycin. DK101 is the progenitor of the A-signal

defective mutants DK4312 (ref. 7) (*asgB* mutation) and MS2021 (ref. 10) (*asgE* mutation), both of which are resistant to kanamycin. All strains are stored as clones in an ultralow freezer.

Sporulation assay

Cultures growing vegetatively in CTT liquid¹⁶ were harvested by centrifugation at 4,900g for 15 min at room temperature and resuspended in 0.5 ml TPM liquid³⁰ at $\sim 5 \times 10^9$ cells ml⁻¹. Resuspended cultures were mixed at appropriate ratios, and 100 μ l of the mixtures were spotted on TPM agar plates. Development progressed at 32 °C for 68 h, when cells were collected into 0.5 ml of sterile H₂O, heated for 2 h at 50 °C, sonicated by microtip, diluted and plated by mixing samples with 10 ml CTT soft agar (0.5% agar). Collected samples containing rifampicin-resistant evolved clones were also plated in CTT soft agar containing 5 μ g ml⁻¹ rifampicin. Spore counts of the rifampicin-resistant genotypes were obtained from plates containing rifampicin, and total spore counts from mixed populations were obtained from plates without antibiotic; spore counts of the rifampicin-sensitive wild-type strains were then calculated as the difference between these plate counts. The plating efficiencies of the rifampicin-resistant genotypes were similar on selective and non-selective agar. The kanamycin- and oxytetracycline-resistant genotypes, however, show decreased plating efficiency during germination on selective relative to non-selective plates. Therefore, to estimate the marked genotype frequency for mixed cultures containing DK4312, MS2021 or LS523, numerous single colonies were transferred (4–5 days after plating) from the non-selective plates to CTT plates containing the appropriate antibiotic (40 μ g ml⁻¹ kanamycin or 12.5 μ g ml⁻¹ oxytetracycline) and the frequency of colonies growing on selective plates was determined after three additional days. Sporulation estimates are the mean of log₁₀-transformed spore counts for at least three replicate assays. A marker-control experiment was run for the rifampicin resistance marker using DK1622 and a developmentally proficient, rifampicin-resistant derivative of DK1622. (The resistant clone is the proximate ancestor of the evolved genotypes used in this study⁶.) This marker-control experiment indicated that rifampicin resistance caused a slight disadvantage during development; however, any handicap associated with this or other resistance markers is conservative with respect to the inferences of complementation and cheating among the evolved clones. Sporulation efficiency is calculated as the ratio of a genotype's spore production to its initial vegetative population size. In Fig. 2c, d, a relative sporulation efficiency of 1 means that an evolved clone and its progenitor produce spores with equal efficiency; values higher than 1 indicate that the evolved clone has a higher sporulation efficiency, whereas values less than 1 indicate higher efficiency of the wild-type progenitor, DK1622.

Received 22 November 1999; accepted 17 January 2000.

- Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
- Axelrod, R. M. *The Evolution of Cooperation* (Basic Books, New York, 1984).
- Pellmyr, O., Leebens-Mack, J. & Huth, C. J. Non-mutualistic yucca moths and their evolutionary consequences. *Nature* **380**, 155–156 (1996).
- Matapurkar, A. K. & Watve, M. G. Altruist cheater dynamics in *Dictyostelium*: aggregated distribution gives stable oscillations. *Am. Nat.* **150**, 790–797 (1997).
- Frank, S. A. *Foundations of Social Evolution* (Princeton Univ. Press, Princeton, 1998).
- Velicer, G. J., Kroos, L. & Lenski, R. E. Loss of social behaviors by *Myxococcus xanthus* during evolution in an unstructured habitat. *Proc. Natl Acad. Sci. USA* **95**, 12376–12380 (1998).
- Kuspa, A., Kroos, L. & Kaiser, D. Intercellular signaling is required for developmental gene expression in *Myxococcus xanthus*. *Dev. Biol.* **117**, 267–276 (1986).
- Shimkets, L. J. & Asher, S. J. Use of recombination techniques to examine the structure of the *csg* locus of *Myxococcus xanthus*. *Mol. Gen. Genet.* **211**, 63–71 (1988).
- Kuspa, A. & Kaiser, D. Genes required for developmental signaling in *Myxococcus xanthus*: three *asg* loci. *J. Bacteriol.* **171**, 2762–2772 (1989).
- Garza, A. G., Harris, B. Z., Pollack, J. S., Tzeng, L. & Singer, M. The *asgE* locus is required for cell-cell signaling during *Myxococcus xanthus* development. *Mol. Microbiol.* **35**, 812–824 (2000).
- Bonner, J. T. *The Cellular Slime Molds* (Princeton Univ. Press, Princeton, 1967).
- Kaiser, D. Control of multicellular development: *Dictyostelium* and *Myxococcus*. *Annu. Rev. Genet.* **20**, 539–566 (1986).
- Reichenbach, H. in *Myxobacteria II* (eds Dworkin, M. & Kaiser, D.) 13–62 (Am. Soc. Microbiol., Washington, DC, 1993).
- Dworkin, M. & Kaiser, D. (eds) *Myxobacteria II* (Am. Soc. Microbiol., Washington, DC, 1993).
- Rosenberg, E. & Varon, M. in *Myxobacteria: Development and Cell Interactions* (ed. Rosenberg, E.) 109–125 (Springer, New York, 1984).
- Kaiser, D. Social gliding is correlated with the presence of pili in *Myxococcus xanthus*. *Proc. Natl Acad. Sci. USA* **76**, 5952–5956 (1979).
- Hagen, D. C., Bretscher, A. P. & Kaiser, D. Synergism between morphogenetic mutants of *Myxococcus xanthus*. *Dev. Biol.* **64**, 284–296 (1978).
- Downard, J., Ramaswamy, S. V. & Kil, K. Identification of *esg*, a genetic locus involved in cell-cell signaling during *Myxococcus xanthus* development. *J. Bacteriol.* **175**, 7762–7770 (1993).
- Laue, B. E. & Gill, R. E. Using a phase-locked mutant of *Myxococcus xanthus* to study the role of phase variation in development. *J. Bacteriol.* **177**, 4089–4096 (1995).
- Buss, L. W. Somatic cell parasitism and the evolution of somatic tissue compatibility. *Proc. Natl Acad. Sci. USA* **79**, 5337–5341 (1982).
- Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **359**, 826–829 (1992).
- Frank, S. A. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* **377**, 520–522 (1995).
- Maynard Smith, J. & Szathmáry, E. *The Major Transitions in Evolution* (Oxford Univ. Press, New York, 1995).
- Grosberg, R. K. & Strathmann, R. R. One cell, two cell, red cell, blue cell: the persistence of a unicellular stage in multicellular life histories. *Trends Ecol. Evol.* **13**, 112–116 (1998).
- Buss, L. W. Slime molds, ascidians, and the utility of evolutionary theory. *Proc. Natl Acad. Sci. USA* **96**, 8801–8803 (1999).

26. Stoner, D. S., Rinkevich, B. & Weissman, I. L. Heritable germ and somatic cell lineage competitions in chimeric colonial protochordates. *Proc. Natl Acad. Sci. USA* **96**, 9148–9153 (1999).
27. Turner, P. E. & Chao, L. Prisoner's dilemma in an RNA virus. *Nature* **398**, 441–443 (1999).
28. Hodgkin, J. & Kaiser, D. Cell-to-cell stimulation of motility in nonmotile mutants of *M. xanthus*. *Proc. Natl Acad. Sci. USA* **74**, 2938–2942 (1977).
29. Wall, D., Kolenbrander, P. E. & Kaiser, D. The *Myxococcus xanthus* pilQ (sglA) gene encodes a secretin homolog required for type IV pilus biogenesis, social motility and development. *J. Bacteriol.* **181**, 24–33 (1999).
30. Kroos, L., Kuspa, A. & Kaiser, D. A global analysis of developmentally regulated genes in *Myxococcus xanthus*. *Dev. Biol.* **117**, 252–266 (1986).

Acknowledgements

We thank A. Garza, L. Shimkets, M. Singer and J. Strassman for helpful discussion, A. Garza for providing strain MS2021, and N. Hajela and J. Jiang for technical assistance. This research was supported by Michigan State University and an NSF grant to R.E.L.

Correspondence and requests for materials should be addressed to G.J.V.
(e-mail: velicerg@pilot.msu.edu).

A family of candidate taste receptors in human and mouse

Hiroaki Matsunami*, Jean-Pierre Montmayeur* & Linda B. Buck

Howard Hughes Medical Institute, Department of Neurobiology,
Harvard Medical School, Boston, Massachusetts 02115, USA

* These authors contributed equally to this work

The gustatory system of mammals can sense four basic taste qualities, bitter, sweet, salty and sour, as well as umami, the taste of glutamate^{1–6}. Previous studies suggested that the detection of bitter and sweet tastants by taste receptor cells in the mouth is likely to involve G-protein-coupled receptors^{2,7,8}. Although two putative G-protein-coupled bitter/sweet taste receptors have been identified⁹, the chemical diversity of bitter and sweet compounds leads one to expect that there is a larger number of different receptors^{8,10,11}. Here we report the identification of a family of candidate taste receptors (the TRBs) that are members of the G-protein-coupled receptor superfamily and that are specifically expressed by taste receptor cells. A cluster of genes encoding human TRBs is located adjacent to a Prp gene locus¹², which in mouse is tightly linked to the SOA genetic locus that is involved in detecting the bitter compound sucrose octaacetate^{13–15}. Another TRB gene is found on a human contig assigned to chromosome 5p15, the location of a genetic locus (PROP) that controls the detection of the bitter compound 6-*n*-propyl-2-thiouracil in humans^{16,17}.

To search for taste receptors, we devised a strategy that was based on four ideas: first, taste receptors would be encoded by a family of related genes; second, some taste receptor genes would be found at genetic loci associated with the ability to taste specific compounds in mouse or human; third, taste receptors would be G-protein-coupled receptors (GPCRs) that have limited sequence similarity to other members of the GPCR superfamily; and last, taste receptor genes might be found by using the resources of the Human Genome Project to look for GPCR-encoding genes in genomic regions implicated in taste perception.

We first asked whether there are genes encoding new GPCRs in the region of the human genome corresponding to the mouse SOA locus, which is tightly linked to a Prp gene^{13,14}. Using the Jackson Laboratory Mouse Genome Informatics website (<http://www.informatics.jax.org>), we determined that the Prp gene and the SOA locus are located on mouse chromosome 6 (63.6 centimorgans (cM)) and that the syntenic region in human is

on chromosome 12p13. We then determined whether any of the genes that mapped close to the mouse SOA locus had been cloned in human and were deposited in the National Center for Biotechnology Information (NCBI) nr database (<http://www.ncbi.nlm.nih.gov>); we used the sequences of those genes to search the NCBI Human Genome Sequence (HGS) database, focusing on human chromosome 12. Among the genes we used for this search was a Prp gene, which we found on the chromosome-12 contig NT_001856. By examining the contig map of chromosome 12 in the NCBI HGS database, we were able to identify contigs that flanked NT_001856. This provided us with a focus set of contigs that might contain taste receptor genes.

To find genes encoding GPCRs in this focus region, we first searched the human chromosome 12 database with large sets of GPCR protein sequences that we compiled from a GPCR database (<http://www.gpcr.org/7tm>). Although we identified a few genes on human chromosome 12 that appeared to encode GPCRs, none was located in the focus set of contigs or their vicinity. We then searched the database with a member of the V1R family of candidate pheromone receptors (V1R5)^{18,19}, because members of this family are not in the GPCR database and therefore had not been included in our GPCR sequence sets. We identified two sequence stretches in contig NT_001856 (which contains the Prp gene) encoding protein sequences distantly related to V1R5. When we retrieved the DNA sequence in and around one of these DNA regions and translated it, we determined that it contained an intronless gene encoding a putative receptor protein of 311 amino acids (hTRB2 (for human taste receptor, family B, no. 2) with weak homology to V1R5 (Fig. 1).

We then asked whether hTRB2 belongs to a family of related receptors, as we expected would be the case for taste receptors. Using hTRB2 to search the chromosome 12 database, we identified eight related genes, all in the NT_001856 contig. Of the eight TRB genes, six encode receptors related to hTRB2 (Fig. 1) and two are pseudogenes. Using these TRBs, we were unable to find any members of this family in either the NCBI nr or expressed sequence tag databases, consistent with the idea that these receptors might be expressed exclusively in taste tissue. However, we did identify a gene encoding a TRB family member (hTRB7, Fig. 1) in a contig assigned to human chromosome 5p15, the location of PROP, the genetic locus that governs the ability of humans to taste 6-*n*-propyl-2-thiouracil, a bitter compound¹⁶. We also found a total of five TRB genes (one a pseudogene) on three chromosome-7 contigs, two of which are assigned to 7q31-32 (data not shown).

The candidate receptors encoded by the TRB genes on chromosomes 12 and 5 (and 7) share sequence motifs with one another, uniting them as members of the same receptor family. Although they have the seven-transmembrane domain structure characteristic of GPCRs, they are unrelated in sequence to both mGluR4, which detects glutamate^{20,21}, and the candidate taste receptors TR1 and TR2 (ref. 9). In addition, mGluR4, TR1 and TR2 have long extracellular amino-terminal domains that are proposed to bind ligand, whereas TRBs have very short N termini, suggesting that they use a different mode of ligand binding. Although TRBs are distantly related to V1Rs, TR1 and TR2 resemble V2Rs (refs 22–24), candidate pheromone receptors that are expressed, with V1Rs, in the vomeronasal organ. The TRBs that we have identified show high variability in protein sequence, suggesting that, like odorant receptors in the nose²⁵, different family members may recognize chemicals with very different structures, such as chemically diverse bitter tastants.

Are TRBs expressed in taste receptor cells, as they must be if they are truly taste receptors? To address this question, we turned to the mouse. We first asked whether we could isolate sequences encoding TRBs from either mouse genomic DNA or complementary DNA prepared from mouse taste tissue. We used polymerase chain reaction (PCR) with degenerate primers matching conserved sequences in TRBs to amplify related sequences, and then cloned

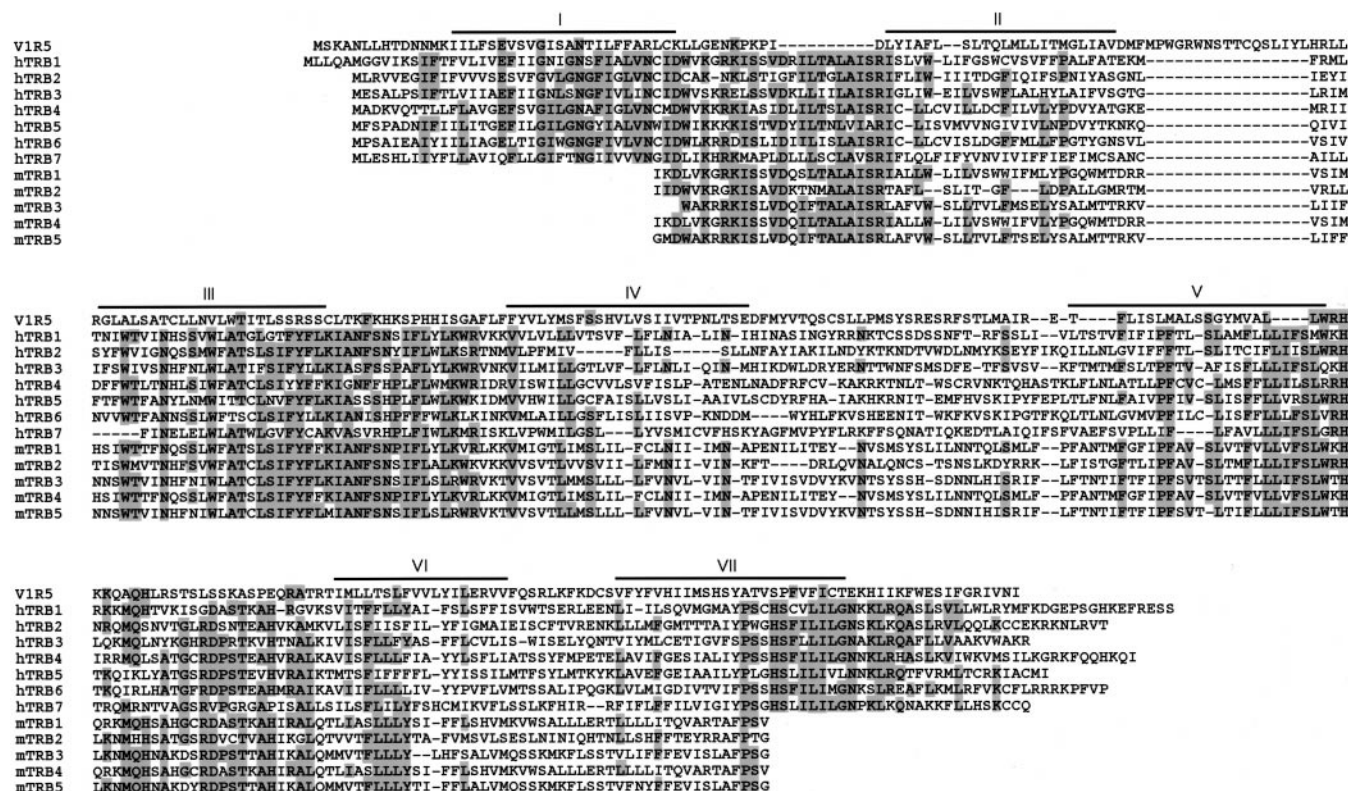


Figure 1 Candidate taste receptors encoded by human and mouse TRB genes. Putative receptor proteins encoded by TRB genes located on human chromosomes 12p13 (hTRB1–6) and 5p15 (hTRB7) or by TRB coding segments amplified from mouse genomic DNA (mTRB1–3) or taste papillae cDNA (mTRB4–5) are aligned here with the putative

pheromone receptor V1R5. Grey bars indicate amino acids found in 50% or more of the TRB proteins shown. Potential transmembrane regions are indicated by horizontal bars and roman numerals (I–VII). The TRBs share amino-acid sequence motifs, but are variable in sequence, consistent with an ability to recognize chemically diverse tastants.

and sequenced the PCR products. Using this approach, we obtained five gene segments encoding TRBs from mouse genomic DNA (Fig. 1), two of which appeared to be derived from pseudogenes (data not shown).

We also obtained three human TRB cDNAs from circumvallate and foliate taste papillae, two of which contained an uninterrupted open reading frame (Fig. 1). These papillae are located on the tongue and contain taste buds where taste receptor cells are clustered¹². These cDNAs were nearly identical in sequence to three of the genes obtained from the genomic DNA. Given the error frequency of the polymerase used for PCR, these differences are unlikely to be due to PCR artifacts. Because the genomic DNA and cDNA were obtained from different mouse strains (DBA/2J and C57BL/6J, respectively), the nearly identical genes could either be closely related or allelic variants.

The presence of a cluster of human genes encoding GPCRs adjacent to a Prp gene, which in mouse cosegregates with SOA, suggested that the TRBs may be taste receptors. Although the mouse and human TRBs are related, none of the mouse TRBs is similar enough to a human TRB identified here to be considered an orthologue. Nonetheless, chromosome mapping of one mouse TRB gene (mTRB3) showed that it is located on mouse chromosome 6 between Prp/D6Mit13 (63.6 cM) and Kap/D6Mit111 (63.7 cM). This corresponds well with the SOA locus, which is tightly linked to the Prp gene¹².

Reverse transcription–PCR (RT–PCR) with either degenerate TRB primers or primers specific for two mTRBs amplified cDNAs of the correct size with RNA from mouse circumvallate and foliate taste papillae as well the tongue, which contains fungiform taste papillae, and the palate, which also has taste buds¹². Appropriately

sized PCR products were also obtained from testis, but not from a number of other tissues tested (Fig. 2). This suggests that TRB genes may be expressed predominantly in taste tissue, as one might predict for genes encoding taste receptors.

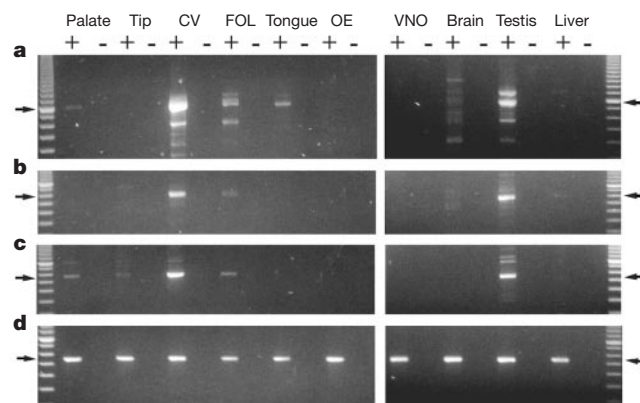


Figure 2 TRB genes are expressed in taste papillae. RT–PCR was carried out with degenerate primers matching human TRBs (a) or primers specific for mTRB3 (b), mTRB2 (c) or mouse β -actin (d). The templates were cDNA prepared from RNA isolated from palate, tongue tip (the location of fungiform taste papillae) (Tip), circumvallate (CV) or foliate (FOL) taste papillae, tongue, olfactory epithelium (OE), vomeronasal organ (VNO), brain, testis or liver. PCR products were fractionated by size on agarose gels with 100-base pair (bp) ladder markers (far left and right lanes). The expected sizes of PCR products are indicated by arrows (a, ~780–800 bp; b, 584 bp; c, 556 bp; d, 540 bp).

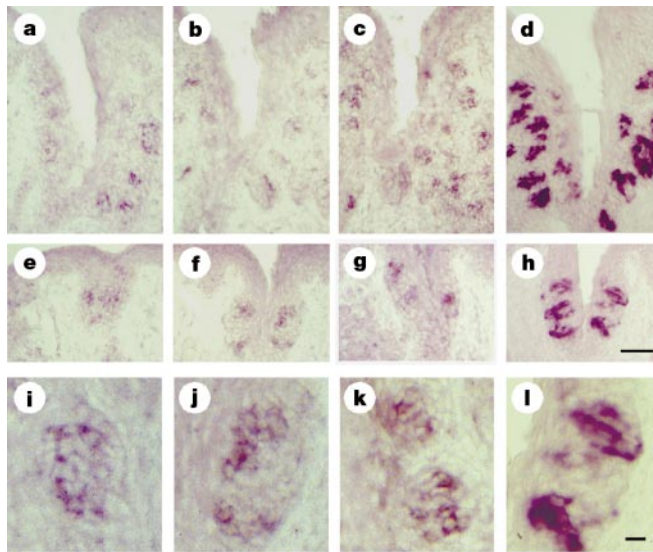


Figure 3 TRB genes are expressed in taste receptor cells in taste buds. Digoxigenin-labelled TRB cRNA probes or a gustducin cRNA probe were hybridized to sections through mouse circumvallate (a–d, i, k) or foliate (e–h, j, l) papillae. The probes used were prepared from sequences encoding mTRB2 (a, e, i), mTRB1 (b, f, j), three different TRBs (mTRB1, mTRB3, mTRB6) (c, g, k) or gustducin (d, h, l). Scale bars, 50 μ m (a–h); shown in h); 10 μ m (i–l; shown in l).

In situ hybridization experiments further showed that TRB genes are expressed in taste receptor cells (Fig. 3). Consistent with the results of the RT–PCR experiments, labelled cRNA receptor probes hybridized to taste receptor cells in the taste buds of both circumvallate and foliate papillae. Moreover, no hybridization was observed to other cells in the taste tissue sections or in sections through the vomeronasal organ, olfactory epithelium or brain. These results indicate that the receptor family that we have identified is probably selectively expressed in taste receptor cells. Notably, the TRB probes that we used labelled only a proportion of taste cells, while a probe matching the gene encoding gustducin, a G protein expressed in ~30–40% of taste receptor cells^{26,27}, labelled a greater number of taste receptor cells (Fig. 3). A mix of TRB probes appeared to label a higher percentage of taste receptor cells than did one probe in the mix, suggesting that different taste receptor cells may express different TRBs (Fig. 3). It may be that other mouse TRB genes not yet identified are expressed either singly or in combinations in other taste receptor cells, and that this allows different taste receptor cells to recognize different tastants.

In summary, we have identified a family of receptors, the TRBs, that are specifically expressed in taste receptor cells in the mouth that detect gustatory stimuli. The TRBs are diverse in protein sequence, which is consistent with an ability to detect a variety of tastants with very different chemical structures. As only a proportion of the human genome sequence has been published, we expect that there will be many additional TRB genes beyond the 11 that we identified on chromosomes 12, 5 and 7. Given that previous studies have implicated G-protein signalling mechanisms in both bitter and sweet taste transduction, members of the TRB family might recognize either bitter or sweet compounds. However, the presence of at least one mouse TRB gene at the SOA locus, and the identification of a cluster of TRB genes at the corresponding site in the human genome suggest that members of this family are likely to be involved in bitter taste sensation. □

Methods

Identification of mouse genes and cDNAs encoding TRBs

RNA was isolated from taste-bud-containing tissues dissected from C57BL/6J mice and then used to prepare cDNA²⁸. PCR reactions were carried out using these cDNAs or genomic DNA from DBA/2J mice (Jackson Laboratory) as described^{22,25} or using a 'touch-down' PCR protocol²⁹. Two pairs of degenerate primers matching human TRBs were used: U1/U8 or U2/U8 (U1, C/TTTG/AG/CIAAT/CGGTTT/CATIGT/C/GIC/GTGTGTA; U2, GGTTT/CATIGT/C/GIC/GTGTGTAAT/CT/GGIATGGA; U8, TT/GA/GTTTICCIAG/TIATAGIAT/A/CIA/CAIGAA/GTG). PCR products were analysed on agarose gels and then gel-isolated and cloned into either pST Blue-1 (Novagen) or pCR4-TOPO (Invitrogen).

Analysis of mouse TRB gene expression

RNA was isolated from taste tissue and from other mouse tissues and used to prepare cDNA as above. The cDNAs were used in PCR reactions with the U1/U8 degenerate primer pair (see above) or with primers specific for mTRB3 (TGTGTGGTCACTACTCAC and GATTGCATGACAAGTGCC), mTRB2 (GAAAGATCTCTGCAGTGGA and GCAAGCCTTTTATGTGGGC) or β -actin (Clontech). PCR was done as described²² except that the degenerate primers were annealed at 40 °C and the specific primers at 55 °C. For *in situ* hybridization experiments, digoxigenin cRNA probes were prepared from cloned mouse TRB gene or cDNA segments and hybridized to 16- μ m tissue sections as described²² except that hybridization was done at 58 °C in a chamber saturated with 50% formamide and 5 $\times$ sodium chloride/sodium citrate (SSC)²².

Chromosome mapping of the mouse TRB3 gene

The location of the mTRB3 gene was determined using the T31-RH Mouse Radiation Hybrid mapping kit (Research Genetics). The data obtained were submitted to the Jackson Laboratory for determination of the chromosomal location of the gene.

Received 13 March; accepted 20 March 2000.

- Roper, S. D. The cell biology of vertebrate taste receptors. *Annu. Rev. Neurosci.* **12**, 329–353 (1989).
- Lindemann, B. Taste reception. *Physiol. Rev.* **76**, 718–766 (1996).
- Lindemann, B. A taste for umami. *Nature Neurosci.* **3**, 99–100 (2000).
- Herness, M. S. & Gilbertson, T. A. Cellular mechanisms of taste transduction. *Annu. Rev. Physiol.* **61**, 873–900 (1999).
- Chaudhari, N. & Roper, S. D. Molecular and physiological evidence for glutamate (umami) taste transduction via a G protein-coupled receptor. *Ann. NY Acad. Sci.* **855**, 398–406 (1998).
- Bartoshuk, L. M. & Beauchamp, G. K. Chemical senses. *Annu. Rev. Psychol.* **45**, 419–449 (1994).
- Kinnamon, S. C. & Margolskee, R. F. Mechanisms of taste transduction. *Curr. Opin. Neurobiol.* **6**, 506–513 (1996).
- Spielman, A. L., Huque, T., Whitney, G. & Brand, J. G. The diversity of bitter taste signal transduction mechanisms. *Soc. Gen. Physiol. Series.* **47**, 307–324 (1992).
- Hoon, M. A. et al. Putative mammalian taste receptors: a class of taste-specific GPCRs with distinct topographic selectivity. *Cell* **96**, 541–551 (1999).
- Lindemann, B. Receptor seeks ligand: on the way to cloning the molecular receptors for sweet and bitter taste. *Nature Med.* **5**, 381–382 (1999).
- Smith, D. V. & Margolis, F. L. Taste processing: whetting our appetites. *Curr. Biol.* **9**, 453–455 (1999).
- Azen, E. et al. Clones from the human gene complex coding for salivary proline-rich proteins. *Proc. Natl Acad. Sci. USA* **81**, 5561–5565 (1984).
- Capeless, C. G., Whitney, G. & Azen, E. A. Chromosome mapping of Soa, a gene influencing gustatory sensitivity to sucrose octaacetate in mice. *Behav. Genet.* **6**, 655–663 (1992).
- Lush, I. E., Hornigold, N., King, P. & Stoye, J. P. The genetics of tasting in mice. VII. Glycine revisited, and the chromosomal location of Sac and Soa. *Genet. Res.* **66**, 167–174 (1995).
- Whitney, G. & Harder, D. B. Genetics of bitter perception in mice. *Physiol. Behav.* **56**, 1141–1147 (1994).
- Reed, D. R. et al. Localization of a gene for bitter-taste perception to human chromosome 5p15. *Am. J. Hum. Genet.* **64**, 1478–1480 (1999).
- Bartoshuk, L. M., Duffy, V. B. & Miller, I. J. PTC/PROP tasting: anatomy, psychophysics, and sex effects. *Physiol. Behav.* **56**, 1165–1171 (1994).
- Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195–206 (1995).
- Rodriguez, I., Feinstein, P. & Mombaerts, P. Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system. *Cell* **97**, 199–208 (1999).
- Chaudhari, N. et al. The taste of monosodium glutamate: membrane receptors in taste buds. *J. Neurosci.* **16**, 3817–3826 (1996).
- Chaudhari, N., Landin, A. M. & Roper, S. D. A metabotropic glutamate receptor variant functions as a taste receptor. *Nature Neurosci.* **3**, 113–119 (2000).
- Matsunami, H., Buck, L. B. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90**, 775–784 (1997).
- Herrada, G. & Dulac, C. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90**, 763–773 (1997).
- Ryba, N. J. & Tirindelli, R. A new multigene family of putative pheromone receptors. *Neuron* **19**, 371–379 (1997).
- Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
- Wong, G. T., Ruiz-Avila, L. & Margolskee, R. F. Directing gene expression to gustducin-positive taste receptor cells. *J. Neurosci.* **19**, 5802–5809 (1999).
- McLaughlin, S. K., McKinnon, P. J. & Margolskee, R. F. Gustducin is a taste-cell specific G protein closely related to the transducins. *Nature* **357**, 563–569 (1992).
- Berghard, A. & Buck, L. B. Sensory transduction in vomeronasal neurons: evidence for G α , G α 2, and adenylyl cyclase II as major components of a pheromone signaling cascade. *J. Neurosci.* **16**, 909–918 (1996).

29. Roux, K. H. in *PCR Primer: A Laboratory Manual* (eds Diffenbach, C. W. & Dveksler, G. S.) 53–62 (CSHL Press, Plainview, New York, 1995).

Acknowledgements

We would like to thank E. Kurosawa-Pelletier and C. Gao for expert technical assistance and K. Fowler for help in preparing the manuscript. This work was supported by the Howard Hughes Medical Institute, grants from the National Institutes of Health (to L.B.B.), and fellowship support from the Alice and Joseph Brook Fund (to J.-P. M.). Genbank accession numbers for mouse TRBs are AF24771–AF24775. Human TRB gene sequences are in the NCBI database in contigs NT_001856 (chromosome 12), NT_000146 (chromosome 5) and NT_001612, NT_000322 and NT_002055 (chromosome 7).

Correspondence and requests for materials should be addressed to L.B.B. (e-mail: lbuck@hms.harvard.edu).

High mobility of proteins in the mammalian cell nucleus

Robert D. Phair* & Tom Misteli†

* BioInformatics Services, Rockville, Maryland 20854, USA

† National Cancer Institute, NIH, Bethesda, Maryland 20892, USA

The mammalian cell nucleus contains numerous sub-compartments, which have been implicated in essential processes such as transcription and splicing^{1,2}. The mechanisms by which nuclear compartments are formed and maintained are unclear. More fundamentally, it is not known how proteins move within the cell nucleus. We have measured the kinetic properties of proteins

in the nucleus of living cells using photobleaching techniques. Here we show that proteins involved in diverse nuclear processes move rapidly throughout the entire nucleus. Protein movement is independent of energy, which indicates that proteins may use a passive mechanism of movement. Proteins rapidly associate and dissociate with nuclear compartments. Using kinetic modelling, we determined residence times and steady-state fluxes of molecules in two main nuclear compartments. These data show that many nuclear proteins roam the cell nucleus *in vivo* and that nuclear compartments are the reflection of the steady-state association/dissociation of its 'residents' with the nucleoplasmic space. Our observations have conceptual implications for understanding nuclear architecture and how nuclear processes are organized *in vivo*.

We measured the mobility of the nucleosomal binding protein HMG-17, the pre-mRNA splicing factor SF2/ASF and the rRNA processing protein fibrillarin in the nucleus of living cells. The three proteins are involved in three essential nuclear functions, transcription, pre-mRNA splicing and rRNA processing, respectively, and they represent three general types of distribution patterns that are observed for many nuclear proteins^{1–3} (Fig. 1). HMG-17 binds to nucleosomes and is thought to alter the higher order chromatin structure by modulating the access of transcriptional regulators to DNA⁴. HMG-17 is distributed evenly in foci throughout the nucleoplasm⁵. SF2/ASF is an essential pre-mRNA splicing factor^{6,7}, which is concentrated in nuclear splicing factor compartments but is also present throughout the nucleoplasm⁸. Fibrillarin associates with the U3 small nucleolar RNA and is involved in the processing of rRNA transcripts in the nucleolus^{9,10}. Fibrillarin is found almost exclusively in the nucleolus¹¹. We expressed these proteins as fusions with the autofluorescent green fluorescent protein (GFP) in HeLa or Baby Hamster Kidney cells. All three GFP-fusion proteins colocal-

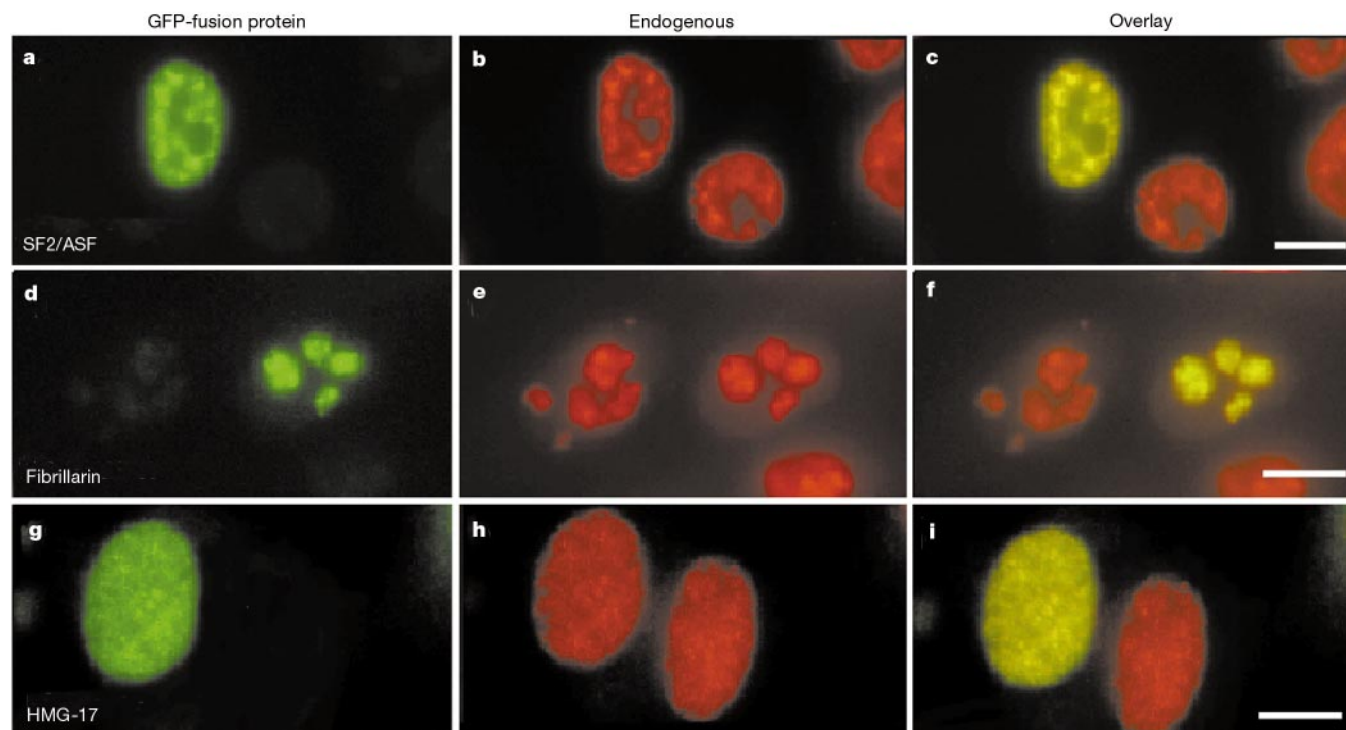


Figure 1 Colocalization of GFP-fusion constructs with endogenous proteins. Fusion proteins between GFP and SF2/ASF (a), fibrillarin (d) and HMG-17 (g) were expressed in BHK cells. Endogenous SF2/ASF (b), fibrillarin (e) and HMG-17 (h) were detected by indirect immunofluorescence using specific antibodies. The distribution pattern of the

GFP-fusion proteins and the endogenous protein were identical and the labelling patterns were identical to that of the endogenous proteins in non-transfected cells (c, f, i). Scale bar, 10 μ m.

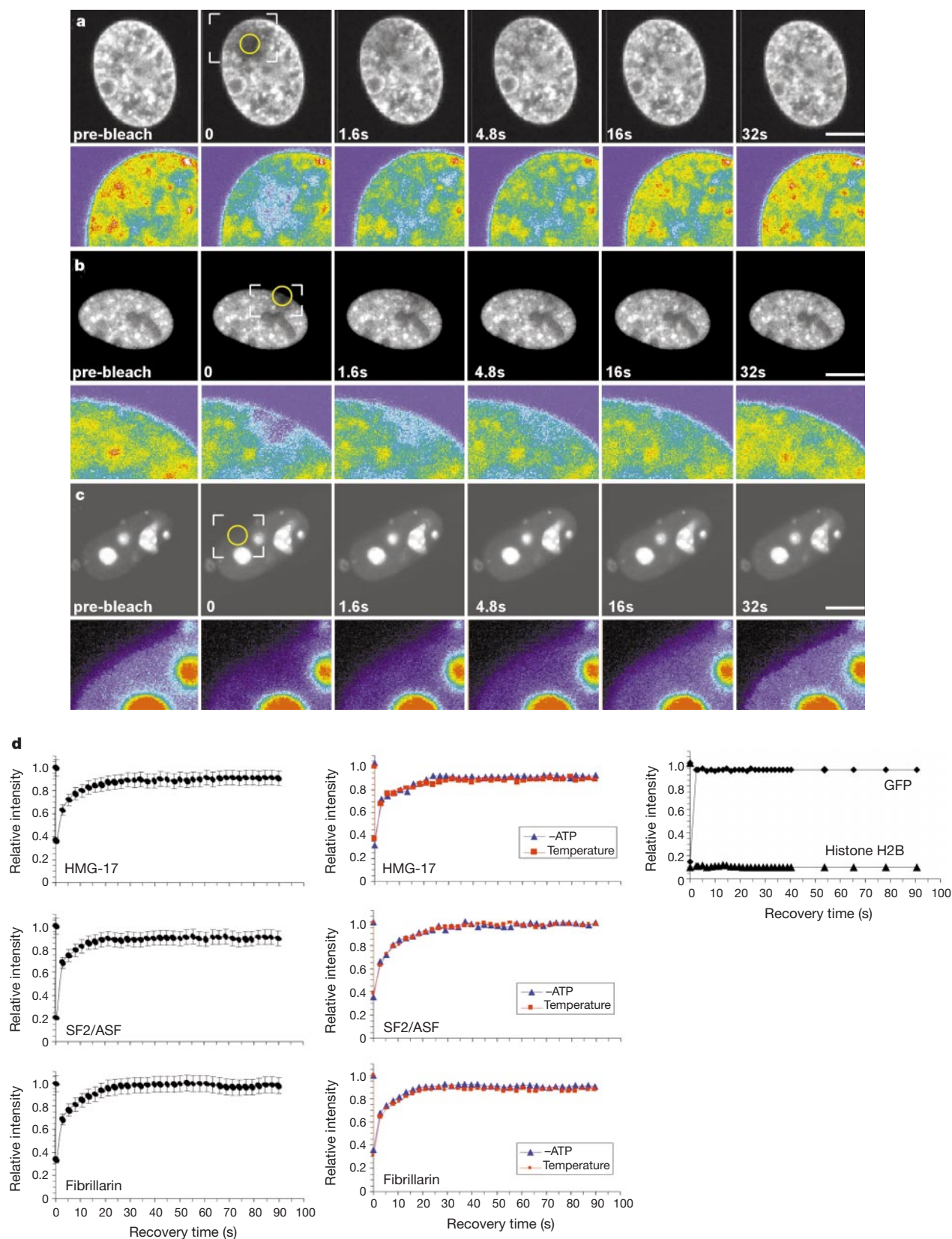


Figure 2 FRAP of nucleoplasmic regions. Cells expressing GFP-HMG-17 (**a**), GFP-SF2/ASF (**b**) or GFP-fibrillarin (**c**) were imaged before and during recovery after bleaching of a nucleoplasmic area for 250 ms. Images were taken at the indicated times after end of the bleach pulse. The area of the bleach spot is indicated with a circle. The indicated area is shown enlarged in pseudocolour in the lower panels. Recovery of all proteins was rapid and was complete within 32 s. Note that the bleached area for fibrillarin is difficult to see owing to the low levels of GFP-fibrillarin in the nucleoplasm. Scale bar, 5 μ m. (**d**) Kinetics

of recovery of nucleoplasmic FRAP. The fluorescence intensity in the bleached region was measured and expressed as the relative recovery (see Methods). Recovery of the fluorescence signal for all three proteins reached a plateau after ~ 30 s, the time for half-recovery was less than 3 s in all cases. Depletion of energy or reduction of temperature to 23 $^{\circ}$ C had no effect on the recovery rate of any of the fusion proteins. As controls, recovery of GFP alone or GFP-histone H2B was measured.

ized with their endogenous counterparts in their expected patterns (Fig. 1)^{5,8,11}. GFP–SF2/ASF has been shown to be functional in alternative splicing *in vivo*¹².

To probe the degree to which nuclear proteins are mobile within the nucleus of a living cell we carried out fluorescence recovery after photobleaching (FRAP)¹³. In our FRAP experiments, a defined area of a living cell is bleached irreversibly by a single, high-powered spot laser pulse of 250 ms. The recovery of the fluorescent signal in the bleached area as the consequence of movement of the GFP-fusion protein is recorded by sequential imaging scans. The kinetics of recovery are a measure for the mobility of the labelled protein¹³. The bleach pulse does not deplete or destroy proteins in the target area¹³. Recovery after photobleaching of a nucleoplasmic area was very fast for all three proteins and the percentage recovery was high suggesting the virtual absence of immobile molecules (Fig. 2). Recovery of GFP–HMG-17 (Fig. 2a), GFP–SF2/ASF (Fig. 2b) and GFP–fibrillarin (Fig. 2c) was complete within 30 s, with a half-time of ~3 s. The diffusion coefficient D in the nucleoplasm for GFP–HMG-17, GFP–SF2/ASF and GFP–fibrillarin was determined from the recovery curves (Fig. 2d; see Methods and Supplementary Information). D values were $0.45 \mu\text{m}^2 \text{s}^{-1}$, $0.24 \mu\text{m}^2 \text{s}^{-1}$ and $0.53 \mu\text{m}^2 \text{s}^{-1}$ for HMG-

17, SF2/ASF and fibrillarin, respectively. These D values are about 100 times lower than values reported for free solutes in the nucleus¹⁴ or for GFP alone^{15,16} (Fig. 2d). As the molecular weight of the fusion proteins is only ~2 times that of GFP, we propose that the lower mobility of the fusion proteins is due to their interaction with other nuclear components. Consistent with these values, the recovery of GFP alone was significantly faster (Fig. 2d). Recovery kinetics of the three proteins were markedly faster than that of GFP–histone H2B (ref. 17), which was immobile over observation periods of up to 4 h (Fig. 2d) presumably because of its tight association with chromatin, which further emphasizes the specificity of the observed mobilities. Although the D values of the fusion proteins are significantly lower than that of GFP alone, the relative movement of the fusion proteins within the nucleus is rapid. Assuming a random walk model, it would take a molecule of GFP–HMG-17 about 28 s to travel a distance of 5 μm , roughly the distance from the centre of the nucleus to the nuclear membrane. The corresponding times for GFP–SF2/ASF and GFP–fibrillarin are 52 s and 24 s, respectively. The movement of all three proteins was independent of energy in the form of ATP and was not significantly slowed at a reduced temperature of 23 °C (Fig. 2d). Both of these results are

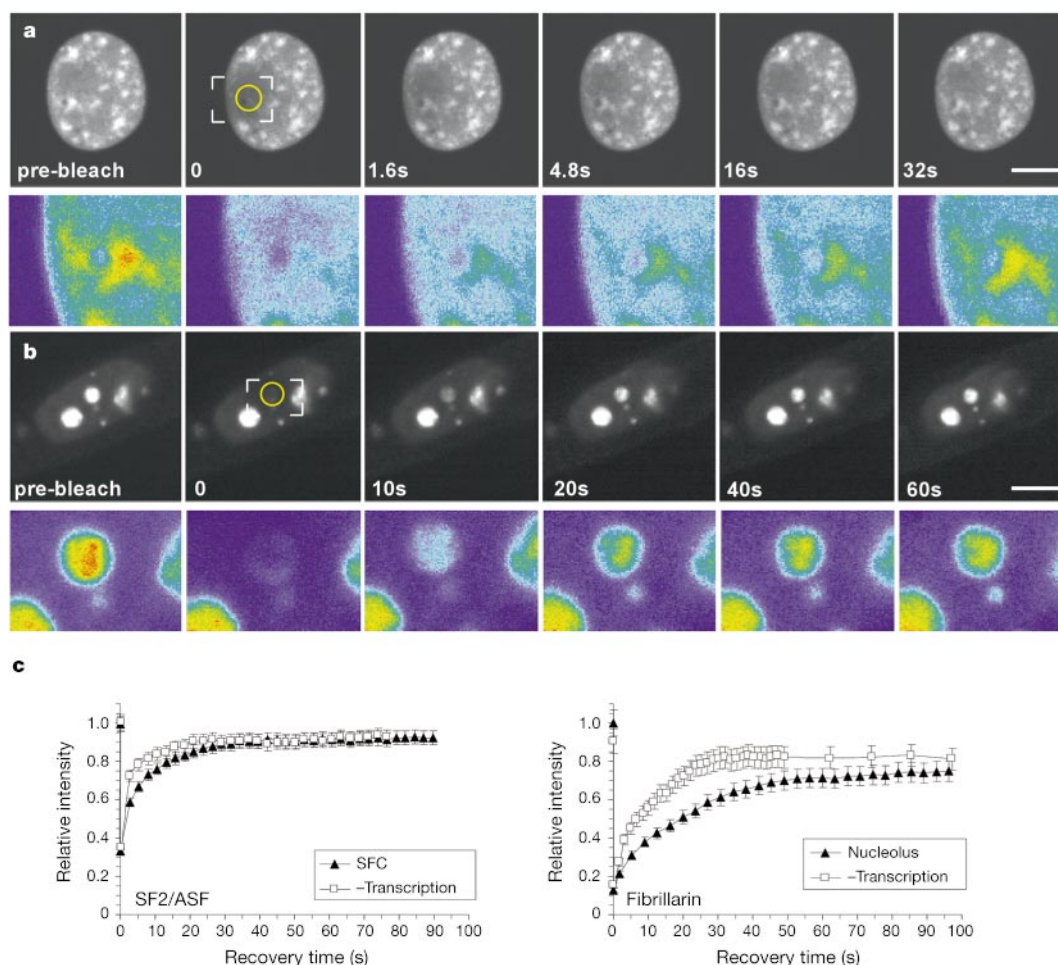


Figure 3 FRAP of nuclear compartments. Cells expressing GFP–SF2/ASF (**a**) or GFP–fibrillarin (**b**) were imaged before and during recovery after bleaching of a splicing factor compartment (**a**) or an entire nucleolus (**b**). Images were taken at the indicated times after the end of the bleach pulse. The area of the bleach spot is indicated with a circle. The indicated area is shown enlarged in pseudocolour in the lower panels. Reassociation of

both proteins with their respective nuclear compartments was rapid. Scale bar, 5 μm . (**c**) Kinetics of recovery after bleaching of nuclear compartments. Recovery of GFP–SF2/ASF was complete after 32 s, recovery of GFP–fibrillarin reached a plateau after 60 s. Inhibition of transcription by either α -amanitin or actinomycin D accelerated the reassociation of GFP–SF2/ASF as well as reassociation of GFP–fibrillarin.

consistent with a diffusion mechanism. Similar results were obtained for GFP-fusion proteins of HGM-14, pre-mRNA splicing factor SC35 and the nucleolar proteins nucleolin and B23 (data not shown). Several control experiments were carried out to confirm the specificity of the FRAP experiments. No recovery of fusion proteins was observed after bleaching of chemically fixed cells and bleaching of cytoplasmic areas did not result in significant loss of fluorescence in the nucleus. Absence of recovery after photobleaching an entire nucleus excluded the possibility that the recovery was the consequence of GFP refolding or *de novo* synthesis. No significant negative effect on cell viability or passage through M phase was observed in cells after bleaching experiments (data not shown). We conclude from these observations that SF2/ASF, HMG-17 and fibrillarin exhibit a high degree of mobility in the nucleoplasm of living cells. The independence of mobility on metabolic energy indicates that proteins in the nucleoplasm may move passively.

Nuclear compartments might represent stable, static protein aggregates; alternatively, proteins might move rapidly in and out of compartments, resulting in steady-state compartments. To distinguish between these possibilities, we selectively bleached single compartments and monitored the recovery of GFP fluorescence. Association of GFP-SF2/ASF and GFP-fibrillarin with the splicing factor compartment or the nucleolus, respectively, was rapid, which

directly shows that these proteins associate at a high rate with their respective compartments (Fig. 3). GFP-SF2/ASF recovery after bleaching of a splicing factor compartment occurred within 30 s (Fig. 3a). An immobile fraction of less than 10% was detected in the splicing factor compartment, suggesting that virtually the entire pool of SF2/ASF in the compartment had been replaced (Fig. 3c). This observation shows that SF2/ASF continuously and rapidly associates with the splicing factor compartment in the nucleus of living cells. After bleaching of an entire nucleolus, GFP-fibrillarin rapidly reassociated from the nucleoplasmic pool (Fig. 3b). The mobility of GFP-fibrillarin within the nucleolus, as measured after bleaching of an area entirely within the nucleolus, was about 10 times lower ($D = 0.046 \mu\text{m}^2 \text{s}^{-1}$) than that of fibrillarin in the nucleoplasm. The association rate of the two proteins with their respective compartments was affected by the transcriptional activity of the nucleus (Fig. 3c). Upon inhibition of transcription of RNA polymerase II by α -amanitin, GFP-SF2/ASF associated slightly, but significantly, faster with the splicing factor compartment (Fig. 3c). Because SF2/ASF predominantly acts on transcripts in the nucleoplasmic space^{1,2}, this increased mobility most probably reflected the absence of nucleoplasmic binding sites in the form of nascent transcripts. In the case of GFP-fibrillarin, inhibition of transcription of RNA pol I by a selectively acting concentration of

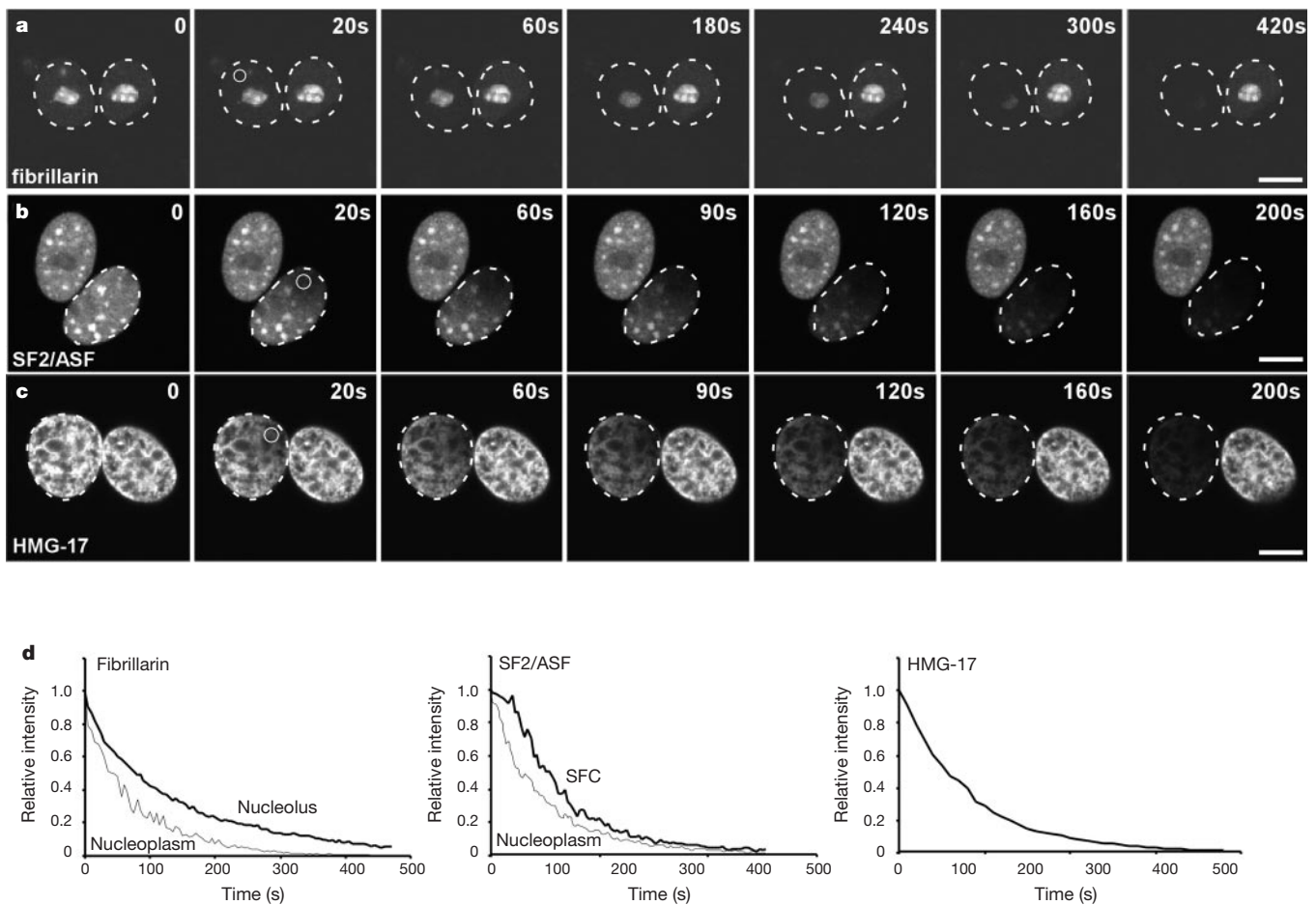


Figure 4 FLIP experiments after bleaching of a nucleoplasmic area. Cells expressing GFP-fibrillarin (**a**), GFP-SF2/ASF (**b**) or GFP-HMG-17 (**c**) were repeatedly bleached for 250 ms in the same nucleoplasmic spot (indicated by a circle). Cells were imaged between bleach pulses. By bleaching a nucleoplasmic area repeatedly, all fluorescence

signal was lost over the indicated period of time, showing a high rate of dissociation of proteins from the nuclear compartments. A neighbouring cell nucleus is not affected by the repeated bleach pulses or the imaging scans. Scale bar, 5 μm . (**d**) Kinetics of loss of fluorescence.

actinomycin D increased the rate of recovery and reduced the immobile fraction of GFP–fibrillarin in the nucleolus (Fig. 3c). The increased mobility is most likely due to the declining level of rRNA transcripts. These observations show that in living cells pre-mRNA splicing factors and nucleolar proteins associate continuously and rapidly with the nuclear compartments in which they are concentrated.

To determine the dissociation kinetics of proteins from compartments, we used fluorescence loss in photobleaching (FLIP)¹³. In FLIP experiments, a living cell is repeatedly bleached in the same spot using high laser power and the cell is imaged before each new round of bleaching. The loss of fluorescence in an area outside of the bleached region, for example, a splicing factor compartment or a nucleolus, is measured over time. The rate of loss of fluorescence from the region of interest contains information on the rate of dissociation of the protein from the particular compartment. By repeatedly bleaching a nucleoplasmic area more than 95% of GFP–fibrillarin fluorescence was lost from all nucleoli (Fig. 4a, d). This observation shows that the nucleolus-associated pool of fibrillarin is continuously and rapidly exchanged with the nucleoplasmic pool. Consistent with this interpretation, repeated bleaching of a nucleolus resulted in the complete loss of GFP–fibrillarin signal in the other nucleoli (data not shown). Similar observations were made for SF2/ASF and HMG-17 (Fig. 4b–d), SC35, HMG-14, nucleolin and B23 (data not shown). The kinetics of loss of signal were consistent with the recovery kinetics observed in FRAP experiments. No significant bleaching was observed in neighbouring nuclei in all of the experiments. No significant loss of fluorescence of the virtually immobile GFP–histone H2B was observed outside the bleached area after 50 bleach pulses and no significant loss of signal was observed in chemically fixed cells, showing that the loss of fluorescence was not due to the formation of quenching agents or by bleaching during the imaging scans (data not shown). These data show that nuclear proteins continuously dissociate from their compartments and that movement of these proteins is not restricted to particular nuclear domains.

We used the FLIP data to develop kinetic models for the distribution and movement of GFP–SF2/ASF and GFP–fibrillarin^{18,19} (see Supplementary Information). Kinetic modelling allowed determination of the relative rate constants for association and dissociation from compartments, the residence time of proteins in compartments and the dissociating flux. The ratio for GFP–SF2/ASF $k_{\text{binding}}/k_{\text{release}}$ was 1.6, indicating that splicing factor compartments concentrate SF2/ASF about 60% more than the nucleoplasm. The corresponding ratio for fibrillarin in the nucleolus was 4.6, reflecting the much higher level of accumulation of fibrillarin in the nucleolus. The maximal mean residence time in the splicing factor compartments was less than 50 s for GFP–SF2/ASF and less than 40 s for fibrillarin in the nucleolus. These values are consistent with the kinetics of recovery measured in FRAP experiments. The residence time of HMG-17 at chromatin cannot be measured using FLIP because it is only present in one morphological pool. However, the almost complete absence of an immobile fraction within 30 s in FRAP experiments (Fig. 2) indicates that the mean residence time is less than 30 s. The minimum number of molecules per second leaving all splicing factor compartments or nucleoli in a single nucleus was ~10,000 molecules of SF2/ASF and ~12,000 molecules of fibrillarin, assuming 10^6 copies of SF2/ASF²⁰ and 5×10^5 molecules of fibrillarin. The short residence times and high rate of movement reinforce the large flux of proteins through the apparently stable nuclear compartments and are a clear indication of the highly dynamic nature of the mammalian cell nucleus.

Our observation that nuclear proteins with diverse functions and distinct distribution patterns move rapidly throughout the interphase cell nucleus has implications for understanding nuclear architecture, morphogenesis of nuclear compartments and nuclear function. The fact that functionally unrelated proteins move at

similar speeds suggests that high mobility is a general feature of nuclear proteins *in vivo*. RNA molecules may move by the same mechanism because the diffusion coefficients of the proteins measured here are similar to those determined for poly(A) RNA and RNP particles^{21–24}. Our results establish that the movement of proteins inside the nucleus of living mammalian cells has kinetic properties that are consistent with a diffusion-based process. The lower mobility of the fusion proteins compared with free GFP probably reflects the interaction, or complex formation, of the proteins with other nuclear components and suggests that putative structural elements within the nucleus might be insufficient to retard the movement of the studied proteins. We find that nuclear proteins continuously and rapidly associate and dissociate with their nuclear compartments. Time lapse fluorescence microscopy has shown that nuclear compartments themselves are stable structures, which persist for long periods of time in a defined nuclear position^{12,25–27}. Our observations that nuclear proteins which ‘reside’ in nuclear structures are nevertheless highly mobile and are rapidly exchanged between the compartments and the nucleoplasm show that these compartments are in continuous flux. This high mobility of proteins suggests that proteins may roam the nucleus rapidly in search of appropriate binding partners or sites. The ability of proteins to cover a large fraction of the nuclear volume in a short period of time raises the possibility that assembly of protein complexes in the nucleus occurs in a stochastic manner and at sites distant from their actual site of action. A stochastic model for formation of protein complexes also maximizes the responsiveness to relatively small changes in the availability of potential binding partners. These considerations indicate that the high mobility of proteins in the cell nucleus make an important contribution to the complex regulatory mechanisms required to coordinate gene expression events *in vivo*. □

Methods

Cell lines and constructs

EGFP–SF2/ASF was generated by inserting SF2/ASF from pGFP–SF2/ASF (ref. 12) into pEGFP–C1 (Clontech) using the *HindIII*, *PstI* restriction sites. EGFP–fibrillarin (a gift from M. Dundr) and EGFP–HMG-17 (a gift from R. Hock) were generated by inserting the full-length complementary DNAs into pEGFP–N3 using the *EcoRI* site or into pEGFP–N2 using the *XhoI* and *EcoRI* sites, respectively. GFP–histone H2B (a gift from G. Wahl) has been described¹⁷. Baby Hamster Kidney cells were transfected by electroporation using a BTX square wave pulser at 650 V, 99 μ s. Transiently transfected cells were observed 14 h after transfection. Indirect immunofluorescence was done as described²⁸ using specific monoclonal antibodies against SF2/ASF (ref. 20) at 1:50, fibrillarin (from Sigma) at 1:1,000 or HMG-17 (a gift from M. Bustin) at 1:50. For inhibition of transcription, cells were incubated for 3 h with either 0.05 μ g ml^{−1} actinomycin D or 25 μ g ml^{−1} α -amanitin. For energy depletion, cells were incubated for 30 min with 6 mM D-glucose and 10 mM NaN₃ (ref. 29).

Microscopy

Cells were plated and observed in LabTek II chambers (Nalgene). Live-cell microscopy was performed on a Leica TCS-SP confocal microscope using the 488 nm laser line of an Ar laser (20 mW nominal output, beam width at specimen 0.2 μ m, detection 500–575 nm). All experiments were done at 37 °C unless otherwise indicated. For FRAP experiments, five single scans were acquired, followed by a single bleach pulse of 200–500 ms using a spot 1 μ m in radius without scanning. Single section images were then collected at 1.6 s intervals. For imaging, the laser power was attenuated to 1% of the bleach intensity. For FLIP experiments, cells were repeatedly imaged and bleached at intervals of either 1 s (SF2/ASF, HMG-17) or 5 s (fibrillarin). Bleaching and imaging settings were identical to the ones used in FRAP. FRAP recovery curves were generated from background subtracted images. The total cellular fluorescence was determined for each image and compared with the initial total fluorescence to determine the amount of signal lost during the bleach pulse and during imaging. In FRAP experiments, typically ~10% of total fluorescence was lost during the bleach pulse and less than 5% of fluorescence was lost during the entire imaging phase. The fluorescence signal measured in a region of interest normalized to the change in total fluorescence was determined as

$$I_{\text{rel}} = \frac{T_0 I_t}{T_t I_0}$$

where T_0 is total cellular intensity during prebleach, T_t the total cellular intensity at timepoint t , I_0 the average intensity in the region of interest during pre-bleach, I_t the

average intensity in the region of interest in timepoint t . This expression accounts for the small loss in total intensity caused by the bleach itself and yields a more accurate estimate of the immobile fraction.

Quantitative kinetic analysis

Nucleoplasmic diffusion coefficients for SF2/ASF, fibrillarin and HMG-17 were determined by classical FRAP analysis³⁰. The data were fitted using the numerical module of the SAAM II software¹⁸ and yielded $K = 3.67$ and $w = 1.16 \mu\text{m}$, where K is the bleach constant and w is the half width of the beam at $1/e^2$ intensity. Coefficients of variation for these values were 1.7% and 2.3%, respectively. The values of K and w were used to fit the fluorescence recovery data to the series solution reported by Axelrod³⁰. To handle large values of K , we truncated the series after 40 terms. This assures that neglected terms make an insignificant contribution to the total. Data recorded for FRAP experiments on SF2/ASF, fibrillarin and HMG-17 in the nucleoplasm of individual nuclei were fitted by a generalized nonlinear least squares technique using the SAAM II software¹⁹. Coefficients of variation were always 10% or less for D and 1% or less for the plateau fluorescence (C_0).

For FLIP analysis of SF2/ASF and HMG-17, we developed kinetic models using SAAM II software (see Supplementary Information).

Received 13 December 1999; accepted 22 February 2000.

1. Lamond, A. I. & Earnshaw, W. C. Structure and function in the nucleus. *Science* **280**, 547–553 (1998).
2. Misteli, T. & Spector, D. L. The cellular organization of gene expression. *Curr. Opin. Cell Biol.* **10**, 322–331 (1998).
3. Matera, A. G. Nuclear bodies: multifaceted subdomains of the interchromatin space. *Trends Cell Biol.* **9**, 302–309 (1999).
4. Bustin, M. Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol. Cell Biol.* **19**, 5237–5246 (1999).
5. Hock, R., Wilde, F., Scheer, U. & Bustin, M. Dynamic relocation of chromosomal protein HMG-17 in the nucleus is dependent on transcriptional activity. *EMBO J.* **17**, 6992–7001 (1998).
6. Ge, H. & Manley, J. L. A protein factor, ASF, controls cell specific alternative splicing of SV40 early pre-mRNA *in vitro*. *Cell* **62**, 24–34 (1990).
7. Krainer, A. R., Conway, G. C. & Kozak, D. The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* **62**, 35–42 (1990).
8. Cáceres, J. F., Misteli, T., Sreteron, G., Spector, D. L. & Krainer, A. R. Role of the modular domains of SR-proteins in subnuclear localization and alternative splicing specificity. *J. Cell Biol.* **138**, 225–238 (1997).
9. Scheer, U. & Hock, R. Structure and function of the nucleolus. *Curr. Opin. Cell Biol.* **11**, 385–390 (1999).
10. Olson, M. O. J., Dundre, M. & Szebeni, A. The nucleolus: An old factory with unexpected capabilities. *Trends Cell Biol.* (in the press).
11. Ochs, R. L., Lischwe, M. A., Spohn, W. H. & Busch, H. Fibrillarin: a new protein of the nucleolus identified by autoimmune sera. *Biol. Cell* **54**, 123–133 (1985).
12. Misteli, T., Cáceres, J. F. & Spector, D. L. The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387**, 523–527 (1997).
13. White, J. & Stelzer, E. Photobleaching GFP reveals protein dynamics inside live cells. *Trends Cell Biol.* **9**, 61–65 (1999).
14. Seksek, O., Biwersi, J. & Verkman, A. S. Translational diffusion of macromolecule-sized solutes in cytoplasm and nucleus. *J. Cell Biol.* **138**, 131–142 (1997).
15. Yokoe, H. & Meyer, T. Spatial dynamics of GFP-tagged proteins investigated by local fluorescence enhancement. *Nature Biotechnol.* **14**, 1252–1256 (1996).
16. Houtsmuller, A. B. *et al.* Action of DNA repair endonuclease ERCC1/XPF in living cells. *Science* **284**, 958–961 (1999).
17. Kanda, T., Sullivan, K. F. & Wahl, G. M. Histone–GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr. Biol.* **26**, 377–385 (1998).
18. Foster, D. M. *et al.* in *Proceedings of the Simulation in Health Science Conference* 87–90 (Society for Computer Simulation, San Diego, 1994).
19. Bell, B. M., Burke, J. V. & Schumitzky, A. A relative weighting method for estimating parameters and variances in multiple data sets. *Computational Statistics and Data Analysis* **22**, 119–135 (1996).
20. Hanamura, A., Cáceres, J. F., Mayeda, A., Franza, B. A. & Krainer, A. R. Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA* **4**, 430–444 (1998).
21. Politz, J. C., Browne, E. S., Wolf, D. E. & Pederson, T. Intracellular diffusion and hybridization state of oligonucleotides measured by fluorescence correlation spectroscopy in living cells. *Proc. Natl Acad. Sci. USA* **95**, 6043–6048 (1998).
22. Daneholt, B. Pre-mRNP particles: From gene to nuclear pore. *Curr. Biol.* **11**, R412–R415 (1999).
23. Politz, J. C., Tuft, R. A., Pederson, T. & Singer, R. H. Movement of nuclear poly(A) RNA throughout the interchromatin space in living cells. *Curr. Biol.* **9**, 285–291 (1999).
24. Singh, O. P., Björkroth, B., Masich, S., Wieslander, L. & Daneholt, B. The intranuclear movement of Balbiani ring premessenger ribonucleoprotein particles. *Exp. Cell Res.* **251**, 135–146 (1999).
25. Sleeman, J. E., Platani, M., Kreiv, J. P., Lamond, A. I. Dynamic interactions between splicing snRNPs, coiled bodies and nucleoli revealed using snRNP protein fusions to the green fluorescent protein. *Exp. Cell Res.* **243**, 290–304 (1998).
26. Boudonck, K., Dolan, L. & Shaw, P. J. The movement of coiled bodies visualized in living plant cells by the green fluorescent protein. *Mol. Biol. Cell* **10**, 2297–2307 (1999).
27. Jolly, C., Usson, Y. & Morimoto, R. I. Rapid and reversible relocation of heat shock factor 1 within seconds to nuclear stress granules. *Proc. Natl Acad. Sci. USA* **96**, 6769–6774 (1999).
28. Misteli, T. & Spector, D. L. Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors. *Mol. Biol. Cell* **7**, 1559–1572 (1996).
29. Dingwall, C., Black, S. J., Kearsey, S. E., Cox, L. S. & Laskey, R. A. Nucleoplasmic cDNA sequence reveals polyglutamic acid tracts and a cluster of sequences homologous to putative nuclear localization signals. *EMBO J.* **6**, 69–74 (1987).
30. Axelrod, D., Koppel, D. E., Schlessinger, J., Elson, E. & Webb, W. W. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophys. J.* **16**, 1055–1069 (1976).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Dundre, G. Wahl, R. Hock and M. Bustin for providing GFP-fibrillarin, GFP-H2B and GFP-HMG-17 clones, respectively, and M. Bustin for anti HMG-17 antibody. We thank M. Bustin for critical reading of the manuscript.

Correspondence and requests for materials should be addressed to T.M. (e-mail: mistelit@mail.nih.gov).

PAR3 is a cofactor for PAR4 activation by thrombin

Mayumi Nakanishi-Matsui*, Yao-Wu Zheng*, David J. Sulciner*†, Ethan J. Weiss*‡, Matthew J. Ludeman* & Shaun R. Coughlin*†‡

* Cardiovascular Research Institute and Daiichi Research Center, † Department of Medicine, and ‡ Department of Cellular and Molecular Pharmacology, University of California, San Francisco, 505 Parnassus Avenue, San Francisco, California 94143-0130, USA

Identification of the mechanisms by which the coagulation protease thrombin activates platelets is critical for understanding haemostasis and thrombosis. Thrombin activates cells at least in part by cleaving protease-activated G-protein-coupled receptors (PARs)¹. PAR3 and PAR4 are thrombin receptors expressed in mouse platelets^{2,3}. Inhibition of thrombin binding to mPAR3 (ref. 4) and knockout of the mPAR3 gene³ inhibited mouse platelet activation at low but not high concentrations of thrombin. Thus PAR3 is important for thrombin signalling in mouse platelets. Expression of human PAR3 in heterologous expression systems reliably resulted in responsiveness to thrombin². Curiously, despite its importance for the activation of mouse platelets by thrombin^{3,4}, mouse PAR3 (mPAR3) did not lead to thrombin signalling even when overexpressed. We now report that mPAR3 and mPAR4 interact in a novel way: mPAR3 does not itself mediate transmembrane signalling but instead functions as a cofactor for the cleavage and activation of mPAR4 by thrombin. This establishes a paradigm for cofactor-assisted PAR activation and for a G-protein-coupled receptor's acting as an accessory molecule to present ligand to another receptor.

We were unable to explain the failure of mPAR3 to result in thrombin signalling on the basis of inability to reach the cell surface, inability to interact with thrombin, cloning artefacts or alternative splicing (not shown). *In situ* hybridization of mouse tissues revealed mPAR3 messenger RNA only in megakaryocytes², raising the possibility that mPAR3 failed to signal in heterologous expression systems because such systems lacked a megakaryocyte/platelet-specific molecule necessary for PAR3 function. Because some G-protein-coupled receptors form heterodimers that are important for function^{5–8}, and because mouse platelets express both mPAR3 and mPAR4 (refs 2, 3), we hypothesized that mPAR3 might interact with mPAR4. Accordingly, we co-expressed mPAR3 and mPAR4 in COS7 cells and examined signalling (Fig. 1). Expression of mPAR3 alone failed to lead to thrombin signalling. mPAR4 alone resulted in thrombin-triggered phosphoinositide hydrolysis with an effector concentration for half-maximum response (EC_{50}) of 0.3 nM. Strikingly, when mPAR3 was expressed with mPAR4, the EC_{50} for thrombin-triggered phosphoinositide hydrolysis decreased to 0.05 nM (Fig. 1). In each of six independent experiments, mPAR3 co-expression decreased the EC_{50} for thrombin signalling by

6–15-fold. This effect was specific for signalling in response to thrombin; mPAR3 co-expression did not enhance signalling in response to the PAR4-activating peptide GYPGKF^{3,9} (data not shown).

Enhanced signalling at low thrombin concentrations in cells co-expressing mPAR3 and mPAR4 might have been because mPAR3 gained the ability to signal when co-expressed with PAR4; that is, mPAR4 might be required for proper biogenesis of mPAR3. Alternatively, mPAR3 cleavage might generate a tethered ligand that ligates mPAR4 *in trans*. In either model, mPAR4 cleavage and

activation would not be required for thrombin signalling in cells co-expressing mPAR3. To test this prediction, we generated the thrombin cleavage site mutant PAR4 G60P. This mutant receptor was expressed on the cell surface and resulted in signalling to GYPGKF but not to thrombin. mPAR3 co-expression did not result in thrombin signalling in cells expressing PAR4 G60P (Fig. 1 and data not shown). Additionally, the synthetic peptide SFNGGP, which mimics the 'tethered ligand domain' of mPAR3 (Figs 2a and 3), failed to cause signalling in COS7 cells expressing mPAR3, mPAR4 or both, even at concentrations as high as 500 μ M (data not shown). These data indicate that mPAR3 does not gain the ability to mediate transmembrane signalling when co-expressed with mPAR4 or ligate mPAR4 *in trans*, but instead promotes signalling through a mechanism that requires mPAR4 cleavage.

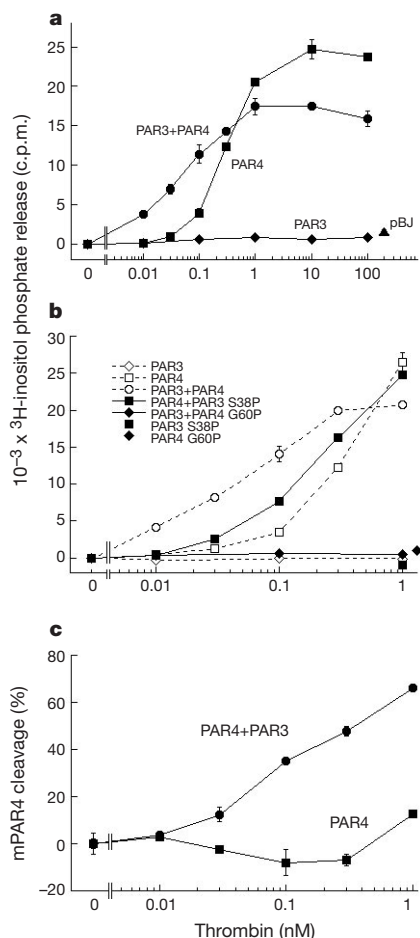


Figure 1 Effect of mPAR3 and mPAR4 co-expression on signalling and receptor cleavage. **a**, ³H-inositol-labelled COS7 cells expressing FLAG-tagged mPAR3 and/or mPAR4 were exposed to the indicated concentration of thrombin together with 10 mM LiCl; phosphoinositide hydrolysis was measured as ³H-inositol phosphates accumulated in 2 h at 37 °C (mean ± s.e.m.; *n* = 2). Similar results were obtained in five independent experiments and when untagged wild-type receptors or shorter incubation times were used. The surface expression of mPAR3 and mPAR4 was measured by the binding of antibody to the FLAG epitope^{13,14}. In cells expressing both receptors, levels of mPAR3 and mPAR4 on the cell surface were similar and were ~70% of those in cells expressing either receptor alone. Thus increased signalling at low thrombin concentrations in mPAR3/mPAR4 cells was not due to an increased surface expression of mPAR4, but the higher maximum signalling in mPAR4-only cells might have been due to higher expression levels. **b**, mPAR3 or mPAR4 cleavage site mutants were expressed alone or together with wild-type receptors; phosphoinositide hydrolysis was measured as in **a**. **c**, Cells were transfected with FLAG-tagged mPAR4 and untagged mPAR3, then incubated as in **a**. Cleavage of mPAR4 was measured as a percentage loss (mean ± s.e.m.; *n* = 2) of the FLAG epitope, which was N-terminal to the thrombin cleavage site, from the cell surface^{13,14}. Similar results were obtained in three replicate experiments and when a shorter incubation time (15 min) was employed.

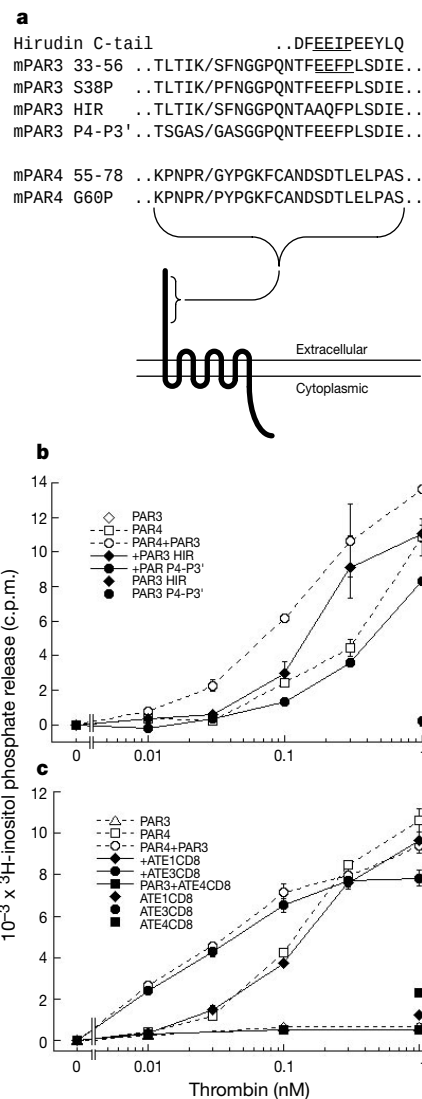


Figure 2 mPAR3's thrombin-interacting sequences are necessary and sufficient to promote mPAR4 activation. **a**, Sequences of the N-terminal exodomains of wild-type and mutant PARs. Note hirudin-like sequence (underlined) carboxy-terminal to the thrombin cleavage site (/) in mPAR3 but not in mPAR4. **b**, **c**, Thrombin-triggered phosphoinositide hydrolysis was measured (see Fig. 1a) in COS7 cells expressing mPAR4 and/or N-terminal exodomain–CD8 fusion proteins (see Methods) (**b**) or mPAR3–thrombin interaction site mutants (**c**) as indicated. Data shown are means ± s.e.m. (*n* = 2). All wild-type and mutant receptors and ATE–CD8 constructs were expressed at similar levels (55–110% of mPAR4). Each experiment was performed at least twice.

mPAR3 is cleaved much more efficiently than mPAR4 (see below), perhaps owing to the hirudin-like sequence in the amino-terminal exodomain of mPAR3 but not that of mPAR4 (Fig. 2). The cognate sequence in hPAR1 is important for thrombin binding and receptor cleavage at low concentrations of thrombin^{10–13}. On the basis of these observations and on the finding that co-expression of mPAR3 and mPAR4 enhanced signalling to thrombin but not to PAR4-activating peptide, we hypothesized that mPAR3 might bind thrombin and promote the cleavage and activation of mPAR4. To test this notion, we co-expressed wild-type mPAR3 with mPAR4 bearing a FLAG epitope N-terminal to the thrombin cleavage site in its N-terminal exodomain³. PAR4 cleavage was followed as a loss of FLAG epitope from the cell surface upon exposure to thrombin^{3,13,14}. Co-expression of mPAR3 caused a striking increase in mPAR4 cleavage at low thrombin concentrations (Fig. 1). The concentration of thrombin required to cleave 50% of cell surface mPAR4 in 15 minutes dropped from 30 nM to only ~1 nM when mPAR3 was co-expressed, whereas the EC₅₀ for the cleavage of mPAR3 itself by thrombin was 0.2 nM (data not shown, but see Fig. 1c). Thus the EC₅₀ for the cleavage of mPAR4 by thrombin in the presence of mPAR3 was intermediate between the EC₅₀ values for the cleavage of mPAR3 and mPAR4 alone.

By analogy with PAR1 (refs 1, 10–13), mPAR3's N-terminal exodomain probably mediates its interaction with thrombin. If mPAR3 acts simply as a binding site for thrombin, the following should hold: (1) displaying mPAR3's N-terminal exodomain on the cell surface should be sufficient to promote mPAR4 activation; (2) the thrombin-interacting sequences in mPAR3's N-terminal exodomain should be necessary for promoting mPAR4 activation; and (3) building mPAR3's thrombin-binding domain into mPAR4 should yield a receptor that is cleaved efficiently at low thrombin concentrations in the absence of mPAR3. Accordingly, we first generated a complementary DNA (cDNA) encoding the N-terminal exodomain of mPAR3 fused to the transmembrane domain of CD8 (ATE3–CD8) so as to display mPAR3's N-terminal exodomain on

the cell surface with the proper topology but lacking the rest of the mPAR3 molecule. When co-expressed with mPAR4 at comparable levels on the cell surface, ATE3–CD8 and mPAR3 were equally active at promoting thrombin signalling, whereas the N-terminal exodomain of mPAR4–CD8 fusion protein was without effect (Fig. 2). The soluble peptide TLTIKSFNGGPQNTFEEFPLSDIEG, which includes mPAR3's thrombin cleavage site and hirudin-like domain, did not promote the cleavage of mPAR4 by thrombin even when added at 1 mM. Thus, when tethered to the cell surface, the N-terminal exodomain of mPAR3 was sufficient to promote mPAR4 activation by thrombin.

We next generated cDNAs encoding mPAR3 mutants in which the P4–P3' residues of mPAR3's thrombin cleavage site or mPAR3's hirudin-like domain were replaced by an irrelevant sequence. Mutation of either site abolished or substantially reduced mPAR3's ability to promote mPAR4 activation (Fig. 2). Thus the thrombin-interacting sequences of mPAR3 are necessary for its ability to promote mPAR4 activation.

Similarly, mutation of the P1' serine in mPAR3's thrombin cleavage site yielded a receptor, designated mPAR3 S38P, that caused only a small enhancement in thrombin responsiveness—less than a two-fold decrease in EC₅₀—when co-expressed with mPAR4 (Fig. 1b). mPAR3 S38P was not cleaved by thrombin (data not shown) and the synthetic peptide TLTIKPFNGGPQNTFEEFPLSDIEG, representing mPAR3's thrombin-interacting sequence

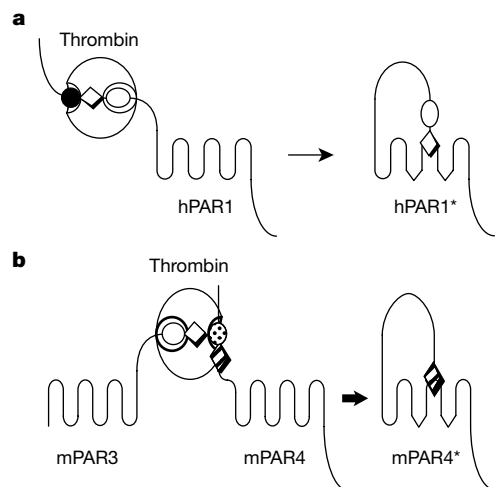


Figure 3 Models of PAR activation. **a**, hPAR1 activation by cleavage *in cis*^{10–13,15,24}. Thrombin (large sphere) binds to the LDPR⁴¹/S⁴²FLLRNPNDKYEPF sequence in hPAR1's N-terminal exodomain. The DKYEPF hirudin-like sequence (oval) binds thrombin's fibrinogen-binding exosite. The P4–P1 LDPR sequence (filled circle) docks with thrombin's active site. Thrombin cleaves the R⁴¹/S⁴² peptide bond, liberating the new N terminus SFLLRN (diamond), which then serves as a tethered ligand that binds intramolecularly to the body of the receptor to trigger transmembrane signalling. **b**, mPAR3 acts as a cofactor for thrombin cleavage and activation of mPAR4, a model of cleavage *in trans*. In one version of this hypothesis, mPAR3 first binds thrombin and is cleaved similarly to hPAR1 in **a**. Thrombin then remains tethered, at least transiently, to the plasma membrane via mPAR3's hirudin-like sequence and cleaves and activates mPAR4.

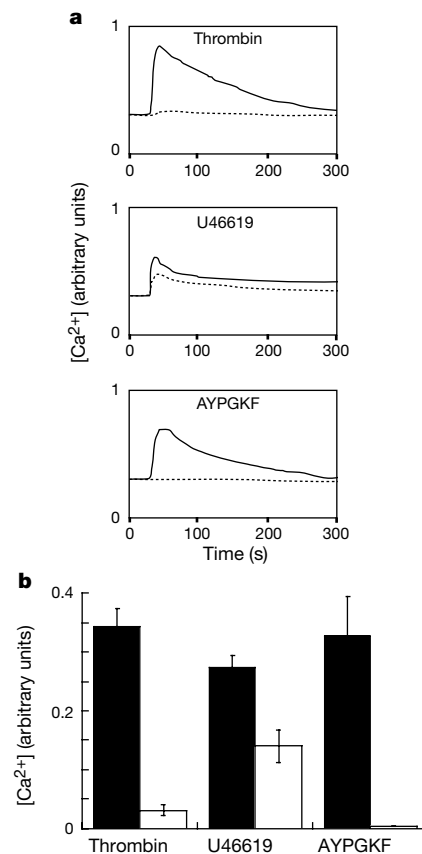


Figure 4 Mouse platelets desensitized with PAR4-activating peptide are refractory to stimulation with thrombin. **a**, Washed Fura-2-loaded mouse platelets were incubated for 30 min at room temperature in the presence (broken lines) or absence (solid lines) of AYPGKF under conditions that inhibited aggregation^{3,9}. Platelets were subsequently stimulated with 'saturating' concentrations of AYPGKF (500 μ M), thrombin (30 nM) or U46619 (50 μ M) and agonist-triggered increases in cytoplasmic Ca²⁺ concentration were measured fluorimetrically. **b**, Peak agonist-triggered Ca²⁺ concentrations (means $\pm$ s.e.m.; $n = 3$) in platelets preincubated in the presence (open) or absence (filled) of AYPGKF were measured as in **a**.

with the S38P substitution, was an effective thrombin inhibitor, blocking cleavage of the small chromogenic substrate S2222 more effectively than the corresponding wild-type mPAR3 peptide (data not shown). It is therefore likely that mPAR3 S38P can bind thrombin but is not cleaved, and that mPAR3 S38P was less active than wild-type mPAR3 in promoting mPAR4 activation because mPAR3 S38P-bound thrombin was effectively inactive, its active site occupied by the uncleavable mPAR3 S38P 'substrate'.

To determine whether the insertion of mPAR3's thrombin-binding sequences into mPAR4 would yield a receptor that is cleaved efficiently at low thrombin concentrations in the absence of mPAR3, we generated mPAR4Hir, a chimera in which mPAR3 residues 44–69 were inserted between mPAR4 residues 65 and 66—in essence, mPAR4 with mPAR3's hirudin-like domain. mPAR4Hir was cleaved even more efficiently than mPAR3 itself. In studies analogous to those in Fig. 1c, the concentrations of thrombin required to cleave 50% of mPAR4Hir, mPAR3 and mPAR4 expressed on the surface of COS7 cells in 15 min were ~0.02, ~0.2 and ~30 nM, respectively.

These data indicate the model shown in Fig. 3. Thrombin binds to mPAR3 through mPAR3's P4–P1 sequence and its hirudin-like domain. Thrombin then cleaves mPAR3, perhaps remaining transiently tethered to the cell surface by mPAR3's hirudin-like sequence. Thrombin so localized, its active centre unoccupied and available to cleave nearby substrates, then cleaves and activates mPAR4. Attempts to address by co-immunoprecipitation the possibility that mPAR3 and mPAR4 form a stable heterodimer have been inconclusive. The observation that ATE3–CD8 was as effective as mPAR3 itself at promoting mPAR4 signalling indicates that heterodimerization, if it occurs, either requires only mPAR3's N-terminal exodomain or is not necessary for mPAR3 to promote mPAR4 activation in the COS7 system (Fig. 2). In the studies depicted in Figs 1 and 2, mPAR3 and mPAR4 were expressed at similar levels on the cell surface and, when surface expression of mPAR3 was decreased fourfold and mPAR4 was held constant, mPAR3's ability to promote mPAR4 activation was virtually abolished (data not shown). Thus mPAR3 must be expressed at levels comparable to those of mPAR4 for significant cofactoring to occur, which is consistent with mPAR3–mPAR4 heterodimerization but also consistent with other models. Thrombin, mPAR3 and mPAR4 might transiently form a ternary complex and/or mPAR3 might become less effective at recruiting thrombin to the cell surface after it is cleaved; both of these models would impose kinetic constraints that might demand comparable levels of mPAR3 and mPAR4 expression.

The studies described above show that mPAR3 can function as a cofactor for mPAR4 cleavage and activation in transfected COS7 cells. However, they do not exclude the possibility that mPAR3 can directly mediate transmembrane signalling in its native environment of the mouse platelet. Again, its failure to signal in heterologous systems might be due to the absence of a platelet-specific factor. If mPAR3 functions only to promote mPAR4 activation and does not itself mediate transmembrane signalling (and if no other thrombin receptors function in mouse platelets), desensitization of mPAR4 in platelets should abolish platelet responsiveness to thrombin. Conversely, if, in 'real platelets', mPAR3 mediates transmembrane signalling, desensitization of mPAR4 should have little effect. To test these predictions, we used a better-than-native PAR4 peptide agonist AYPGKF. At the concentrations used, this peptide elicited robust increases in cytoplasmic Ca^{2+} concentrations in cells expressing human or mouse PAR4 but not PAR1, PAR2 or PAR3 (Faruqi, T. R., E.J.W., Huang, W. and S.R.C., manuscript in preparation). Mouse platelets preincubated with AYPGKF were refractory to subsequent challenge with AYPGKF, and thrombin-triggered increases in cytosolic Ca^{2+} concentration were virtually ablated (8% of those in undesensitized platelets) (Fig. 4). By contrast, responses to the thromboxane A_2 receptor agonist

U46619 were relatively preserved (52% of undesensitized) (Fig. 4). Importantly, human platelets express PAR1 and PAR4, and both receptors mediate transmembrane signalling^{3,9,15,16}. In contrast with mouse platelets, the desensitization of human platelets with AYPGKF did not decrease thrombin responsiveness (ref. 9 and data not shown). Thus if mPAR3 mediated transmembrane signalling in mouse platelets like hPAR1 in human platelets, thrombin signalling should be preserved after desensitization of mouse platelets with AYPGKF. The lack of thrombin signalling after desensitization with AYPGKF indicates that PAR4 is necessary for all or virtually all thrombin signalling in mouse platelets, which is consistent with the model in Fig. 3. Together with the observation that the loss of mPAR3 function inhibited mouse platelet activation by low concentrations of thrombin^{3,4}, these results indicate that the physiological role of mPAR3 in mouse platelets is that of a cofactor for mPAR4 activation. This model makes the testable prediction that, unless as yet undescribed thrombin receptors contribute, thrombin signalling should be ablated in platelets from PAR4-deficient mice, now being generated.

Because human platelets express PAR1 and PAR4 but apparently little or no PAR3 (ref. 9), and because hPAR1 and hPAR4 can each mediate thrombin signalling independently^{3,9,16}, the mouse system just described will not find a direct analogy in the human platelet. However, mPAR3 did promote cleavage and activation of hPAR4 as effectively as it did that of mPAR4 (data not shown); thus, hPAR4 can be 'cofactored'. hPAR1 and the GPIIb α –GPIIb β –GPIX complex, a platelet thrombin-binding site¹⁷, are obvious candidates for hPAR4 cofactors in human platelets, but so far the co-expression of these proteins with hPAR4 in COS7 cells has not enhanced hPAR4 cleavage at low thrombin concentrations. Experiments with PAR1-inhibited human platelets indicate that relatively high thrombin concentrations are required to activate hPAR4 in this cell type⁹. Thus, it remains unknown whether or not cofactors have a direct role in human PAR4 activation. Nevertheless, the paradigm of cofactor-dependent PAR activation raises important new questions and will probably be open to generalization in other ways (see below).

In summary, our studies show that mPAR3 is a cofactor for mPAR4 cleavage and activation by thrombin. This is an intriguing example of evolution of molecular function and reveals remarkable opportunism. In one species, a PAR has 'evolved' from signalling receptor to cofactor by losing transmembrane signalling but not thrombin recognition. The latter function assumed or retained importance for mouse platelet function by promoting mPAR4 activation at low thrombin concentrations. Whether or in what circumstances other PARs exhibit cofactor activity or benefit from the actions of cofactors requires exploration. The coagulation cascade is replete with cofactors that bind proteases to the cell surface and facilitate their interaction with substrates¹⁸, and it seems likely that such cofactors might facilitate PAR activation in certain settings. Indeed, the expression of tissue factor, the factor VIIa cofactor, allows PAR2-expressing cells to respond to picomolar factor VIIa in a factor X-dependent manner (E. Camerer and S.R.C., manuscript in preparation). The use of cofactors in PAR activation might promote specificity by demanding that the target cell express both cofactor and PAR to be responsive to a particular protease. Regulation of cofactor expression might also permit cells to vary their responsiveness to distinct proteases while using only one or a few PARs. More generally, the paradigm of one receptor's acquiring a ligand that in turn triggers transmembrane signalling through a distinct receptor is well established in some systems^{19–23} but to our knowledge has not been reported between G-protein-coupled receptors. PAR3's cofactor function in mouse platelets shows that G-protein-coupled receptors can function as accessory molecules that promote the ligation of other G-protein-coupled receptors. G-protein-coupled receptors are known to interact in several ways. For example, heterodimerization of GABA $_B$ R1 and

GABA_BR2 is required for the formation of a functional GABA_B receptor^{5,7,8}, and the heterodimerization of κ and δ opioid receptors results in a receptor with distinct ligand-binding specificity⁶. The cofactor relationship between mPAR3 and mPAR4 adds to this increasingly rich repertoire of interactions. □

Methods

Constructs and transfections

COS7 cells were transfected with pBJ1-based expression vectors (Mark Davis, Stanford) by using Lipofectamine (Gibco). Each transfection included 6 μ g DNA per 8 $\times 10^5$ cells per 6 cm plate (3 μ g each of wild-type or mutant mPAR3 and/or mPAR4 vector or empty expression vector). Cells were split into 12-well or 24-well plates; after 48 h, phosphoinositide hydrolysis, receptor expression and cleavage were measured^{13,14}. Except when stated, expressed proteins had a prolactin signal peptide joined to the FLAG epitope at their N termini. mPAR3 cDNA thus tagged encoded the N terminus MDSKGSS-QKGSRLLLLVSNLLCQGVVS/DYKDDDDVET... representing the prolactin signal peptide, signal peptidase site (/), the FLAG epitope DYKDDDD and junction VE fused to amino acid Thr 17 in mPAR3. Tagged mPAR4 contained the same sequence as far as DDDDV fused to amino acid Ala 18 of mPAR4. ATE3-CD8 encoded a type 1 single transmembrane domain protein in which the epitope-tagged N-terminal exodomain of PAR3 as far as amino acid Thr 93 was fused to Ile 162 at the extracellular end of CD8's single transmembrane domain. ATE4-CD8 encoded the epitope-tagged N-terminal exodomain of mPAR4 as far as amino acid Ser 79 fused to Ile 162 in CD8 (ref. 24).

Assays

Phosphoinositide hydrolysis in transfected COS7 cells was measured as described¹⁴. Agonist-triggered increases in intracellular Ca²⁺ concentration were measured fluorometrically in Fura-2-labelled washed platelets from C57/BL6 mice as described^{2,3,25}. For mPAR4 desensitization, platelets were incubated with 500 μ M AYPGKF for 30 min at room temperature without being stirred in platelet buffer⁴ containing 0.1 mM EGTA. Studies of inhibition of chromogenic substrate cleavage by thrombin with the use of PAR peptides were performed as described¹¹.

Received 21 December 1999; accepted 31 January 2000.

- Coughlin, S. R. How the protease thrombin talks to cells. *Proc. Natl Acad. Sci. USA* **96**, 11023–11027 (1999).
- Ishihara, H. *et al.* Protease-activated receptor 3 is a second thrombin receptor in humans. *Nature* **386**, 502–506 (1997).
- Kahn, M. L. *et al.* A dual thrombin receptor system for platelet activation. *Nature* **394**, 690–694 (1998).
- Ishihara, H., Zeng, D., Connolly, A. J., Tam, C. & Coughlin, S. R. Antibodies to protease-activated receptor 3 inhibit activation of mouse platelets by thrombin. *Blood* **91**, 4152–4157 (1998).
- Kuner, R. *et al.* Role of heteromer formation in GABA_B receptor function. *Science* **283**, 74–77 (1999).
- Jordan, B. A. & Devi, L. A. G-protein-coupled receptor heterodimerization modulates receptor function. *Nature* **399**, 697–699 (1999).
- White, J. H. *et al.* Heterodimerization is required for the formation of a functional GABA_B receptor. *Nature* **396**, 679–682 (1998).
- Kaupmann, K. *et al.* GABA_B-receptor subtypes assemble into functional heteromeric complexes. *Nature* **396**, 683–687 (1998).
- Kahn, M. L., Nakanishi-Matsui, M., Shapiro, M. J., Ishihara, H. & Coughlin, S. R. Protease-activated receptors 1 and 4 mediate activation of human platelets by thrombin. *J. Clin. Invest.* **103**, 879–887 (1999).
- Vu, T. -K. H., Wheaton, V. I., Hung, D. T. & Coughlin, S. R. Domains specifying thrombin–receptor interaction. *Nature* **353**, 674–677 (1991).
- Liu, L., Vu, T. -K. H., Esmon, C. T. & Coughlin, S. R. The region of the thrombin receptor resembling hirudin binds to thrombin and alters enzyme specificity. *J. Biol. Chem.* **266**, 16977–16980 (1991).
- Mathews, I. I. *et al.* Crystallographic structures of thrombin complexed with thrombin receptor peptides: existence of expected and novel binding modes. *Biochemistry* **33**, 3266–3279 (1994).
- Ishii, K., Gerszten, R., Zheng, Y. -W., Turck, C. W. & Coughlin, S. R. Determinants of thrombin receptor cleavage: receptor domains involved, specificity, and role of the P3 aspartate. *J. Biol. Chem.* **270**, 16435–16440 (1995).
- Ishii, K., Hein, L., Kobilka, B. & Coughlin, S. R. Kinetics of thrombin receptor cleavage on intact cells. Relation to signaling. *J. Biol. Chem.* **268**, 9780–9786 (1993).
- Vu, T. -K. H., Hung, D. T., Wheaton, V. I. & Coughlin, S. R. Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* **64**, 1057–1068 (1991).
- Xu, W. F. *et al.* Cloning and characterization of human protease-activated receptor 4. *Proc. Natl Acad. Sci. USA* **95**, 6642–6646 (1998).
- Okamura, T., Hasitz, M. & Jamieson, G. A. Platelet glycolalcalin: interaction with thrombin and role as thrombin receptor on the platelet surface. *J. Biol. Chem.* **253**, 3435–3443 (1978).
- Colman, R. W., Marder, V. J., Salzman, E. W. & Hirsh, J. in *Hemostasis and Thrombosis* (eds Colman, R. W., Marder, V. J., Salzman, E. W. & Hirsh, J.) 3–18 (Lippincott, Philadelphia, 1994).
- Ulevitch, R. J. & Tobias, P. S. Receptor-dependent mechanisms of cell stimulation by bacterial endotoxin. *Annu. Rev. Immunol.* **13**, 437–457 (1995).
- Chow, J. C., Young, D. W., Golenbock, D. T., Christ, W. J. & Gusovsky, F. Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *J. Biol. Chem.* **274**, 10689–10692 (1999).
- Heinrich, P. C., Behrmann, I., Müller-Newen, G., Schaper, F. & Graeve, L. Interleukin-6-type cytokine signalling through the gp130/Jak/STAT pathway. *Biochem. J.* **334**, 297–314 (1998).
- Klagsbrun, M. Mediators of angiogenesis: the biological significance of basic fibroblast growth factor (bFGF)–heparin and heparan sulfate interactions. *Semin. Cancer Biol.* **3**, 81–87 (1992).

- Vivien, D., Attisano, L., Wrana, J. L. & Massagué, J. Signaling activity of homologous and heterologous transforming growth factor- β receptor kinase complexes. *J. Biol. Chem.* **270**, 7134–7141 (1995).
- Chen, J., Ishii, M., Wang, L., Ishii, K. & Coughlin, S. R. Thrombin receptor activation: confirmation of the intramolecular tethered ligand hypothesis and discovery of an alternative intermolecular liganding mode. *J. Biol. Chem.* **269**, 16041–16045 (1994).
- Sage, S. O. in *Platelets, A Practical Approach* (eds Watson, S. P. & Authi, K. S.) 67–90 (IRL, Oxford, 1996).

Acknowledgements

We thank H. Bourne, T. Nakanishi, M. Shapiro and J. Trejo for critical reading of the manuscript.

Correspondence and requests for materials should be addressed to S.R.C. (e-mail: coughlin@cvrmail.ucsf.edu).

ATM phosphorylates p95/nbs1 in an S-phase checkpoint pathway

Dae-Sik Lim^{*†}, Seong-Tae Kim^{*†}, Bo Xu^{*}, Richard S. Maser[‡], Junyu Lin[‡], John H.J. Petrini[‡] & Michael B. Kastan^{*}

^{*} Department of Hematology–Oncology, St. Jude Children's Research Hospital, 332 North Lauderdale Street, Memphis, Tennessee 38105, USA

[‡] Laboratory of Genetics, University of Wisconsin, Genetic Building, 445 Henry Mall, Madison, Wisconsin 53706, USA

[†] These authors contributed equally to this work

The rare diseases ataxia-telangiectasia (AT), caused by mutations in the ATM gene, and Nijmegen breakage syndrome (NBS), with mutations in the p95/nbs1 gene, share a variety of phenotypic abnormalities such as chromosomal instability, radiation sensitivity and defects in cell-cycle checkpoints in response to ionizing radiation^{1–4}. The ATM gene encodes a protein kinase that is activated by ionizing radiation or radiomimetic drugs^{5,6}, whereas p95/nbs1 is part of a protein complex that is involved in responses to DNA double-strand breaks^{3,7}. Here, because of the similarities between AT and NBS, we evaluated the functional interactions between ATM and p95/nbs1. Activation of the ATM kinase by ionizing radiation and induction of ATM-dependent responses in NBS cells indicated that p95/nbs1 may not be required for signalling to ATM after ionizing radiation. However, p95/nbs1 was phosphorylated on serine 343 in an ATM-dependent manner *in vitro* and *in vivo* after ionizing radiation. A p95/nbs1 construct mutated at the ATM phosphorylation site abrogated an S-phase checkpoint induced by ionizing radiation in normal cells and failed to compensate for this functional deficiency in NBS cells. These observations link ATM and p95/nbs1 in a common signalling pathway and provide an explanation for phenotypic similarities in these two diseases.

Optimal induction and activation of p53 protein after ionizing radiation requires activation of the ATM kinase^{5,6}. AT cells show suboptimal increases in p53, defects in phosphorylation on Ser 15 and poor arrest at stage G1 of the cell cycle after ionizing radiation^{8–11}. In contrast, there have been conflicting reports on the nature of p53 induction and G1 arrest in NBS cells^{12,13}. To determine whether p95/nbs1 is required for ATM activation in response to ionizing radiation, we examined steps in the ATM–p53 pathway in NBS cells. The specific activity of ATM kinase increased roughly twofold after ionizing irradiation in lymphoblastoid cells derived from both NBS patients and normal individuals (Fig. 1a). A control AT cell line exhibited no detectable ATM activity (data not shown). Similarly, increases in p53, phosphorylation of p53 on Ser 15 and p53-dependent induction of p21 and G1 arrest were not detectably different in lymphoblasts from NBS patients compared with those

from normal individuals (Fig. 1b and data not shown). Thus p95/nbs1 is not required for the activation of ATM and its downstream targets after ionizing radiation.

We then investigated whether p95/nbs1 is a physiological downstream target of ATM. Irradiation of cells resulted in slower migration of p95/nbs1 protein on SDS–polyacrylamide gel electrophoresis (SDS–PAGE) analysis (Fig. 2a, lanes 2 and 6), indicating that p95/nbs1 was modified in response to DNA damage. This slower migration was eliminated by treatment of the immunoprecipitated p95/nbs1 with a serine/threonine phosphatase, but not after treatment with a tyrosine phosphatase (Fig. 2a, lanes 4 and 5). This ionizing radiation-induced alteration in p95/nbs1 migration failed to occur in AT cells (Fig. 2b, lane 7). Phosphorylation of p95/nbs1 was still not evident in AT cells 6 h after 5 Gy ionizing radiation, indicating that it is absent, not simply delayed (Fig. 2c). A small alteration in p95/nbs1 migration can be seen in AT cells following higher doses of ionizing radiation at later time points (data not shown). AT cells complemented with an ATM complementary DNA showed restored phosphorylation of p95/nbs1 induced by ionizing radiation (Fig. 2d). In contrast to dependence on ATM, a normal ionizing radiation-induced p95/nbs1 band shift in DNA-dependent protein kinase catalytic subunit (DNA-PKcs)-null cells indicates that DNA-PKcs is not required for ionizing radiation-induced phosphorylation of p95/nbs1 *in vivo* (Fig. 2d). These results indicate that ionizing radiation may induce a serine/threonine phosphorylation event on p95/nbs1 that is dependent on ATM, and raise the possibility that p95/nbs1 may be a direct physiological substrate for ATM *in vivo* after ionizing radiation.

To investigate whether ATM directly phosphorylates p95/nbs1 after ionizing radiation, we first evaluated the ability of ATM to phosphorylate p95/nbs1 *in vitro*. ATM belongs to a family of PI3-like kinases that includes DNA-dependent protein kinase (DNA-PK) and ataxia-telangiectasia and rad3-related kinase (ATR)^{14,15}. We previously characterized preferred substrate sequences for ATM, ATR and DNA-PK, and found that ATM has a strong preference for phosphorylating SQ/TQ motifs with hydrophobic amino acids at positions N-3 and N-1 (ref. 16). The amino-acid sequence around Ser 343 of p95/nbs1 contains hydrophobic amino acids at N-3

and N-1 as well as an SQ motif, which is also present in the amino-acid sequence near Ser 15 in p53 (Fig. 3a). A polypeptide containing the 14 amino acids surrounding Ser 343 of p95/nbs1 linked to GST was readily phosphorylated by wild-type, but not kinase-dead, ATM (Fig. 3b). The level of phosphorylation was similar to that seen with the p53 peptide target sequence (Fig. 3b). A much larger glutathione S-transferase (GST)–peptide containing 65 amino acids surrounding Ser 343 of p95/nbs1 (amino acids 327–391) was also an excellent *in vitro* substrate for ATM (Fig. 3c). Mutation of Ser 343 to alanine in the target GST–p95/nbs1 largely abrogated its phosphorylation by ATM, implicating Ser 343 as a target site of *in vitro* phosphorylation. In addition, whereas ionizing radiation increased *in vivo* ³²P metabolic labelling of a transfected full-length p95/nbs1 in an ATM-dependent manner, an S343A mutant of p95/nbs1 exhibited significantly less incorporation of ³²P after ionizing radiation (Fig. 3d). As the S343A p95/nbs1 mutant did exhibit some ³²P incorporation, however, there must be at least one other site in the protein that can be phosphorylated.

To confirm that ATM kinase phosphorylates p95/nbs1 on Ser 343 in response to ionizing radiation *in vivo*, we generated a mouse polyclonal antibody that specifically recognizes p95/nbs1 when it is phosphorylated on Ser 343. The specificity of the antibody was illustrated by its ability to recognize a p95/nbs1 peptide sequence surrounding Ser 343 in its phosphorylated state, but not its unphosphorylated state (Fig. 4a). Specificity was also confirmed by a failure of the antibody to bind to p95/nbs1 from unirradiated cells or to the p95/nbs1 S343A mutant (Fig. 4b). Use of this antibody showed that ionizing radiation induces the *in vivo* phosphorylation of Ser 343 in a p95/nbs1 protein that had been exogenously introduced into 293T cells (Fig. 4b, lanes 2). The endogenous ATM protein in these 293T cells is wild type and can be induced by ionizing radiation (data not shown). To examine the ATM dependence of this ionizing radiation-induced phosphorylation event, either wild-type or kinase-inactive ATM was co-transfected into 293T cells along with the epitope-tagged p95/nbs1 vector. The basal level of Ser 343

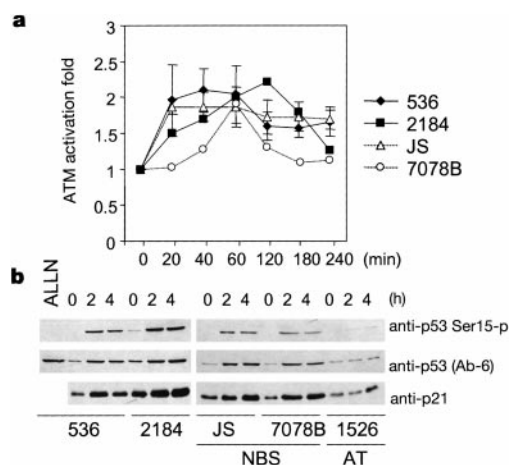


Figure 1 Activation of ATM kinase and the p53 pathway by ionizing radiation occurs in NBS cells. **a**, Relative increase in ATM activity compared with unirradiated controls. Normal (GM0536, GM02184) or NBS (JS, GM07078B) lymphoblasts were either untreated or treated with 5 Gy ionizing radiation and collected at the indicated times after irradiation, and the specific activity of ATM was quantified as described⁹. **b**, Immunoblot analysis of p53 protein levels (middle), phosphorylation of p53 on Ser 15 (top) and p21 protein levels (bottom) at various times after ionizing radiation in normal, NBS and AT cells. Normal cells treated for 2 h with the proteasome inhibitor ALLN served as a control for the Ser 15 antibody.

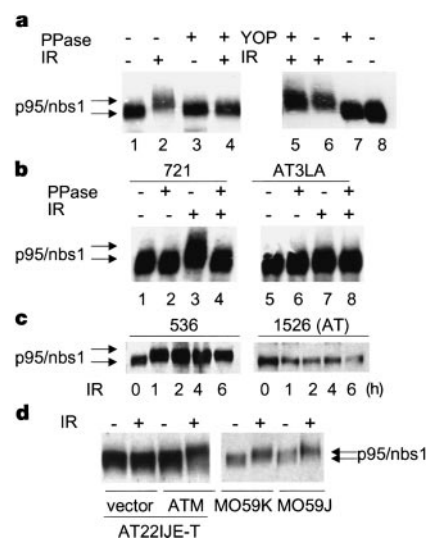


Figure 2 Ionizing radiation (IR) induces ATM-dependent serine/threonine phosphorylation of p95/nbs1. **a**, Normal lymphoblasts (721) were treated with either 0 (–) or 10 (+) Gy irradiation and p95/nbs1 was immunoprecipitated, either untreated (–) or treated (+) with lambda phosphatase (PPase) or tyrosine phosphatase (YOP), and analysed by SDS–PAGE and western blotting for p95. The top arrow indicates the slower mobility band. **b**, Mobility shift of p95/nbs1 after ionizing radiation in normal (721) versus AT (AT3LA) lymphoblasts. **c**, Mobility shift of p95/nbs1 at various times after 5 Gy irradiation in normal (GM0536) versus AT (GM01526) lymphoblasts. **d**, Mobility shift of p95/nbs1 at 45 min after 10 Gy irradiation in AT212JE-T fibroblasts expressing either empty vector or full-length ATM²⁷ or in DNA-PK-proficient cells and -deficient cells (MO59K and MO59J, respectively).

phosphorylation in unirradiated cells after co-transfection with wild-type ATM was slightly higher than that seen with vector controls, probably reflecting a partial activation of overexpressed ATM (Fig. 4b, lanes 1 and 3). However, there was still a significant increase in phosphorylation of p95/nbs1 Ser 343 after ionizing radiation (Fig. 4b, lanes 2 and 4). In contrast, co-transfection of the kinase-inactive form of ATM significantly inhibited this ionizing radiation-induced phosphorylation of p95/nbs1 Ser 343 (Fig. 4b, lanes 5 and 6). This inhibition of ionizing radiation-induced phosphorylation of Ser 343 reflects the ability of this kinase-inactive form of ATM to function in a dominant-negative manner when overexpressed. These results show that ionizing radiation treatment of cells results in phosphorylation of Ser 343 in p95, the same site shown to be an ATM target *in vitro*, and that this ionizing radiation-induced phosphorylation is blocked by use of a dominant-negative form of ATM.

As a final step in demonstrating that ATM kinase is required for the phosphorylation of p95/nbs1 Ser 343 induced by ionizing radiation, we also examined the phosphorylation of Ser 343 in endogenous p95/nbs1 in response to ionizing radiation in normal and AT cells. Endogenous p95/nbs1 was immunoprecipitated from unirradiated or irradiated lymphoblasts from a normal individual or an AT patient and subsequently probed for phosphorylation with the anti-p95/nbs1 phosphoserine 343 antibody. Phosphorylation of Ser 343 of p95/nbs1 was easily detected in irradiated cells containing wild-type ATM (Fig. 4c, lane 2), but no phosphorylation was observed in either irradiated or unirradiated cells from two AT patients (Fig. 4c, lanes 3–6). Introduction of ATM into AT cells restored this ionizing radiation-induced phosphorylation (Fig. 4c, lane 8). Equivalent amounts of total immunoprecipitated p95/nbs1 were present under all conditions. These results show that p95/nbs1 protein is phosphorylated on Ser 343 in an ATM-dependent manner

in response to ionizing radiation.

We then explored the functional consequences of ionizing radiation-induced phosphorylation of p95/nbs1 Ser 343 by ATM. p95/nbs1 forms a complex with hRad50 and hMre11, and this complex has been implicated in the recognition and repair of DNA double-stranded breaks^{1,2,17–19}. The formation of hMre11–hRad50 foci induced by ionizing radiation is defective in NBS cells and reduced in SV-40-transformed AT cells^{1,20}, although there is a much less pronounced effect of ATM deficiency on focus formation in primary fibroblasts²¹. To test whether ATM-dependent p95/nbs1 phosphorylation on Ser 343 affects the formation of foci, we transfected either wild-type or S343A mutant haemagglutinin (HA)-tagged p95/nbs1 into HeLa cells and examined foci induced by ionizing radiation. Indirect immunofluorescence analysis showed a minimal decrease in the formation of ionizing radiation-induced foci with the S343A mutant in HeLa cells (data not shown). In addition, retroviral stable expression of either wild-type or the S343A mutant of p95/nbs1 in NBS primary fibroblasts restored the formation of ionizing radiation-induced hMre11/hRad50 foci (data not shown). We also found that the S343A p95/nbs1 bound normally to hMre11 (Fig. 5a). Thus, ATM phosphorylation of p95/nbs1 does not appear directly to affect the binding of p95/nbs1 to hMre11 or have a major impact on the formation of ionizing radiation-induced foci.

As loss of the ionizing radiation-induced S-phase checkpoint (also commonly referred to as radioresistant DNA synthesis (RDS)) is a common phenotype of both AT and NBS cells⁴, we assessed the transient inhibition of DNA synthesis after exposure to ionizing radiation in cells expressing wild-type or the S343A mutant of p95/nbs1. At all time points tested, DNA synthesis in 293T cells transiently transfected with vector or wild-type p95/nbs1 exhibited ionizing radiation-induced S-phase arrest (Fig. 5b). In contrast,

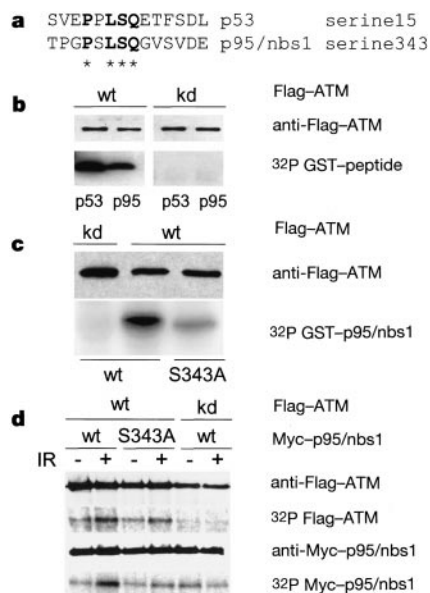


Figure 3 Phosphorylation of p95/nbs1 by ATM. **a**, Amino-acid sequence surrounding p53 Ser 15 and p95/nbs1 Ser 343. **b**, Phosphorylation of GST–p53 (amino acids 9–22) and GST–p95/nbs1 peptide (amino acids 337–350) by wild-type (wt) or kinase dead (kd) Flag-tagged ATM. Flag-tagged ATM in each kinase reaction was blotted with anti-Flag (top panel). **c**, GST–p95/nbs1 (327–391) or GST–p95/nbs1 (327–391) with Ser 343 substituted by alanine (S343A) was used as a substrate for ATM in *in vitro* kinase reactions. **d**, 293T cells transfected with wild-type or S343A mutant Myc–p95/nbs1 and wt or kd Flag–ATM were treated with 0 (–) or 5 (+) Gy ionizing radiation and labelled with [³²P]orthophosphate for 30 min. ³²P labelling of immunoprecipitated ATM (second panel) and p95/nbs1 (bottom panel) was analysed by autoradiography and protein levels assessed by immunoblot.

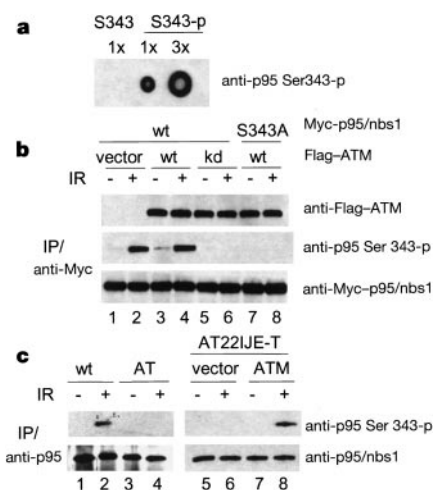


Figure 4 ATM-dependent phosphorylation of p95/nbs1 on Ser 343 *in vivo* after ionizing radiation. **a**, Immunoblot of synthetic p95/nbs1 peptides (S343, unphosphorylated peptide (337–350); S343-p, peptide (337–350) phosphorylated at Ser 343) with mouse polyclonal antibody raised against the S343-p peptide. 1x indicates 50 ng. **b**, 293T cells were co-transfected with wild-type or S343A mutant Myc–p95/nbs1 along with either empty vector or wild-type (wt), or kinase inactive (kd) Flag–ATM. After 48 h, cells were treated with either 0 (–) or 5 (+) Gy ionizing radiation and harvested after 30 min. Flag–ATM in each cell lysate was assessed by immunoblot (top), and immunoprecipitated Myc–p95/nbs1 was blotted with anti-α-p95-phosphoserine 343 antibody (middle) or anti-Myc antibody (bottom). **c**, Normal (wt, GM0536) or AT (GM01526) lymphoblasts, or AT221JE-T fibroblasts expressing empty vector or ATM²⁷ were treated with either 0 (–) or 5 (+) Gy ionizing radiation. Endogenous p95/nbs1 was immunoprecipitated 30 min after ionizing radiation and blotted with anti-α-p95-phosphoserine 343 antibody (top) or anti-p95/nbs1 antibody (bottom).

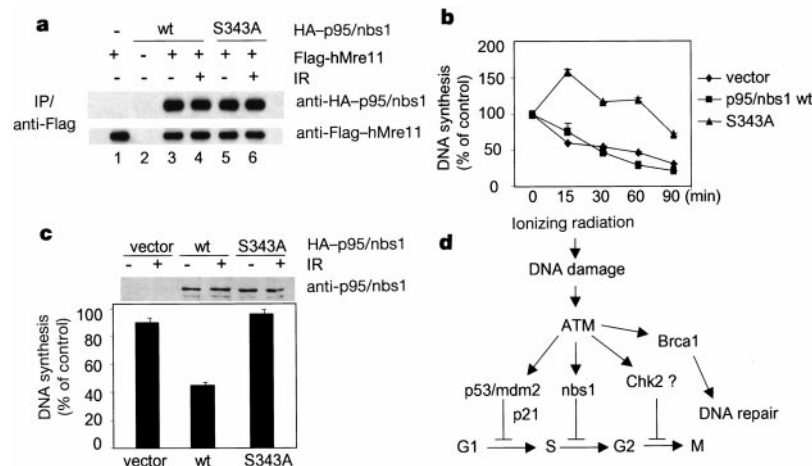


Figure 5 Mutation of the ATM phosphorylation site in p95/nbs1 results in an S-phase checkpoint defect (RDS). **a**, 293T cells were transfected with the indicated plasmids and co-immunoprecipitation of HA-p95/nbs1 with Flag-hMre11 was assessed using the anti-Flag antibody for immunoprecipitation and either anti-HA (top) or anti-Flag (bottom) for immunoblotting. **b**, 293T cells transiently expressing empty vector, wild-type (wt) or S343A mutant p95/nbs1 were assessed for DNA synthesis at various times after

treatment with 20 Gy ionizing radiation. **c**, NBS primary fibroblasts were infected with a retroviral vector expressing empty vector, wild-type (wt) or S343A mutant p95/nbs1 and were assessed for DNA synthesis 30 min after 20 Gy ionizing radiation. Expression of wild-type and S343A p95/nbs1 was verified by immunoblotting (top). **d**, Current understanding of ionizing radiation-induced signalling pathways involving ATM in mammalian cells.

293T cells expressing an S343A mutant showed significantly decreased S-phase arrest. This abrogation of the S-phase checkpoint is similar to that seen with transfection of the kinase-inactive, dominant-negative ATM (data not shown). This observation indicates that the S343A mutant of p95/nbs1 can also act in a dominant-negative fashion in this pathway.

To confirm that the phosphorylation of p95/nbs1 on S343 by ATM is required for the transient inhibition of replication during S phase, we infected NBS primary fibroblasts with retrovirus encoding either wild-type or S343A mutant p95/nbs1. Ectopic expression of the wild-type p95/nbs1 protein complemented the RDS defect and restored the ionizing radiation-induced inhibition of DNA synthesis (Fig. 5c). In contrast, neither the empty vector nor the p95/nbs1 S343A mutant rescued the RDS phenotype of NBS primary fibroblasts (Fig. 5c). This result was also observed in stable polyclonal populations selected for retroviral insertion (data not shown). Expression levels of the wild-type and S343A mutant form of p95/nbs1 protein in these infected cells were similar (Fig. 5c). Thus, cells expressing the mutant form of p95/nbs1 that is resistant to ATM phosphorylation continue to manifest the RDS phenotype seen in both NBS and AT cells. This means that ionizing radiation induces phosphorylation of Ser 343 of p95/nbs1 in an ATM-dependent manner, ATM phosphorylates this same site in p95/nbs1 in *in vitro* kinase assays, and, although this ATM-mediated phosphorylation event does not appear to be required for complex formation with hMre11/hRad50, it is functionally related to the ionizing radiation-induced S-phase checkpoint.

On the basis of these data, we conclude that loss of the ionizing radiation-induced S-phase checkpoint in AT and NBS cells can be explained by the biochemical and functional link between ATM and p95/nbs1 (Fig. 5d). Therefore, we propose a model in which ATM regulates p53 activity to control ionizing radiation-induced G1 arrest by affecting phosphorylation of p53 and mdm2 (refs 5, 6, 22) and controls S-phase arrest after ionizing radiation by phosphorylating p95/nbs1 on Ser 343. The mechanism by which phosphorylation of Ser 343 affects DNA replication after ionizing radiation is unknown. This ATM-dependent modification could either directly affect the function of p95/nbs1 or indirectly influence other functionally important modifications of the protein. ATM also appears to regulate the hChk2 kinase, which may contribute to the inhibition of G2-M transition^{23,24} or to p53 regulation. It has also

been reported that ATM phosphorylates Brca1 (refs 16, 25) and that Brca1 can bind to the nbs1-hMre11-hRad50 complex²⁶. Therefore, understanding the functional relationship between ATM, the nbs1-hMre11-hRad50 complex and Brca1 could provide further insight into the mechanisms underlying abnormal ionizing radiation responses observed in AT, NBS and Brca1-null cells. □

Methods

Antibodies

The anti-Flag M2 (Sigma), anti-c-Myc or anti-HA (Roche), anti-p53 Ab-6 (Calbiochem) and anti-p53-phosphoserine 15 (New England Biolabs) antibodies were used for immunoblotting and immunoprecipitation. The p95/nbs1 rabbit polyclonal antibody has been described¹. We generated the α -p95-phosphoserine 343 polyclonal antibody by immunizing mice with a KLH-conjugated phosphopeptide (TPGPSL(PO₃)SQGVSVDE). The α -p95-phosphoserine 343 antiserum was purified by affinity chromatography using phosphopeptide-conjugated Sepharose, and antibody components directed against non-phosphospecific epitopes were removed on a Sepharose column conjugated with an unphosphorylated peptide (TPGPSLSQGVSVDE). The ATM monoclonal antibodies (D16.11 for immunoprecipitation and D16.35 for immunoblotting) were gifts from T. M. Gilmer.

Plasmids

The GST fusion p53 and p95/nbs1 peptide expression vectors were as described¹⁶. To construct the larger GST-p95/nbs1 (327–391) protein fragment, p95/nbs1 DNA encoding amino acids 327–391 was amplified by PCR with Pfu polymerase using the following primers: 5' sense, 5'-TCCCCAGGAATCCCGGCCATCCAGTACAGGATTA-3'; 3' antisense, 5'-TGCGGCCGCTCGAGTTTGTGTCCATTGTGGAGAC-3'. The amplified PCR product was digested with *EcoRI*-*XhoI* and was cloned into pGEX-4T-2 (Pharmacia). GST-p95/nbs1 (327–391) S343A was generated using the QuikChange Site-Directed Mutagenesis kit (Stratagene). To construct the HA-tagged or Myc-tagged p95/nbs1 expression vector, the entire coding region of p95/nbs1 was PCR-amplified with a pair of oligonucleotides: 5'-GAATCCCTCGAGCTACCGCCATGTGGAACACTGCTG CCGCCGCG-3' and 5'-GTCGACGAGCGGCCGCCACCTCAGGGATCTTCTCC TTTTAAATAAGG-3'. The PCR product was digested with *XhoI*-*NotI* and cloned into a pSG5 vector (Neupogen) that had a HA or c-Myc tag. The HA-tagged p95/nbs1 wild type and S343 mutant were cloned into a pBABE puro²⁷ to generate the retroviral expression vector.

In vitro kinase assays and immunoprecipitation

In vitro kinase assays for Flag-tagged ATM and endogenous ATM were performed as described³. For immunoprecipitation of p95/nbs1, cells were irradiated with 0 or 10 Gy ionizing radiation and lysed 45 min after irradiation in ice-cold lysis buffer (100mM NaCl, 10 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5% Igepal CA630, 10 mM sodium β -glycerophosphate, 5 mM Na₃VO₄, 1 mM PMSF). After centrifugation, supernatants were incubated with either anti-p95/nbs1 antibody or anti-hMre11 antibody. After extensive washing with a lysis buffer, immunoprecipitants were analysed by immunoblot.

For lambda phosphatase treatment, immunoprecipitated p95/nbs1 was washed with a phosphatase buffer (50 mM Tris-HCl pH 7.5, 2 mM MnCl₂, 0.1 mM EDTA, 5 mM DTT, 0.01% Brij 35) and incubated with 400 units of lambda phosphatase (New England Biolabs) at 30 °C for 30 min. For YOP protein tyrosine phosphatase treatment, the immunocomplexes were washed in YOP phosphatase buffer (50 mM Tris-HCl pH 7.0, 100 mM NaCl, 2 mM EDTA, 5 mM DTT, 0.01% Brij 35, 1 mg ml⁻¹ BSA) and incubated with 50 units of YOP (New England Biolabs).

In vivo phosphorylation of p95/nbs1

Myc-tagged p95/nbs1 (3 µg) was transiently co-transfected into 293T cells with wild-type ATM, kinase-dead ATM or pCDNA3 (7 µg) using calcium phosphate. After 2 days, Myc-p95/nbs1 from cell lysates was immunoprecipitated and subsequently immunoblotted with an anti-Myc antibody or an anti-p95-phosphoserine 343 polyclonal mouse antibody. For metabolic labelling, the cells were washed with phosphate-free RPMI plus 10% dialysed fetal bovine serum and incubated for 30 min before irradiation. The cells were irradiated at 5 Gy and incubated with [³²P]orthophosphate (1 mCi ml⁻¹) for 30 min at 37 °C. Radiolabelled p95/nbs1 and ATM were immunoprecipitated and analysed on a PhosphorImager.

RDS assay

Inhibition of DNA synthesis after ionizing radiation was performed as described²⁸. After transfection with Myc-p95/nbs1 (10 µg), 293T cells were labelled with 10 nCi of [¹⁴C]thymidine per ml for 24 h to control for total DNA content between samples. Then cells were incubated with nonradioactive medium for 24 h and treated with 0, 5, 10 or 20 Gy of ionizing radiation, incubated for various times as noted, and then labelled with 2.5 µCi ml⁻¹ of [³H]thymidine for 15 min. After fixing cells, radioactivity was measured in a liquid scintillation counter. The DNA synthesis was measured by the ratios of ³H/¹⁴C and expressed as a percentage of unirradiated cells. No significant changes in [³H]thymidine incorporation were observed in unirradiated cells transfected with the various vectors. For the RDS assay in cells retrovirally infected with HA-p95, viral supernatant derived from 293T cells that had been co-transfected with pBABE p95/nbs1 and pEQPAM3 were infected into NBS cells (GM07166) for 5 h. After infection, cells were labelled with 10 nCi of [¹⁴C]thymidine per ml for 24 h. The medium was replaced with radioactive free medium and incubated for 24 h. Other procedures were done as described above. Two days after infection, cells were selected in the presence of puromycin (2 µg ml⁻¹) for five days. These stable polyclonal puromycin resistant populations were also used for the RDS assay and expression of p95/nbs1 was confirmed by immunoblotting.

Received 20 October 1999; accepted 9 February 2000.

- Carney, J. P. *et al.* The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell* **93**, 477–486 (1998).
- Varon, R. *et al.* Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* **93**, 467–476 (1998).
- Featherstone, C. & Jackson, S. P. DNA repair: the Nijmegen breakage syndrome protein. *Curr. Biol.* **8**, R622–R625 (1998).
- Shiloh, Y. Ataxia-telangiectasia and the Nijmegen breakage syndrome: related disorders but genes apart. *Annu. Rev. Genet.* **31**, 635–662 (1997).
- Canman, C. E. *et al.* Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* **281**, 1677–1679 (1998).
- Banin, S. *et al.* Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science* **281**, 1674–1677 (1998).
- Petrini, J. H. The mammalian Mre11-Rad50-nbs1 protein complex: integration of functions in the cellular DNA-damage response. *Am. J. Hum. Genet.* **64**, 1264–1269 (1999).
- Kastan, M. B. *et al.* A mammalian cell cycle checkpoint pathway utilizing p53 and GADD45 is defective in ataxia-telangiectasia. *Cell* **71**, 587–597 (1992).
- Canman, C. E., Wolff, A. C., Chen, C. Y., Fornace, A. J. Jr & Kastan, M. B. The p53-dependent G1 cell cycle checkpoint pathway and ataxia-telangiectasia. *Cancer Res.* **54**, 5054–5058 (1994).
- Khanna, K. K. *et al.* Nature of G1/S cell cycle checkpoint defect in ataxia-telangiectasia. *Oncogene* **11**, 609–618 (1995).
- Siliciano, J. D. *et al.* DNA damage induces phosphorylation of the amino terminus of p53. *Genes Dev.* **11**, 3471–3481 (1997).
- Jongmans, W. *et al.* Nijmegen breakage syndrome cells fail to induce the p53-mediated DNA damage response following exposure to ionizing radiation. *Mol. Cell. Biol.* **17**, 5016–5022 (1997).
- Yamazaki, V., Wegner, R. D. & Kirchgessner, C. U. Characterization of cell cycle checkpoint responses after ionizing radiation in Nijmegen breakage syndrome cells. *Cancer Res.* **58**, 2316–2322 (1998).
- Hartley, K. O. *et al.* DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell* **82**, 849–856 (1995).
- Cimprich, K. A., Shin, T. B., Keith, C. T. & Schreiber, S. L. cDNA cloning and gene mapping of a candidate human cell cycle checkpoint protein. *Proc. Natl Acad. Sci. USA* **93**, 2850–2855 (1996).
- Kim, S. T., Lim, D. S., Canman, C. E. & Kastan, M. B. Substrate specificities and identification of putative substrates of ATM kinase family members. *J. Biol. Chem.* **274**, 37538–37543 (1999).
- Nelms, B. E., Maser, R. S., MacKay, J. F., Lagally, M. G. & Petrini, J. H. *In situ* visualization of DNA double-strand break repair in human fibroblasts. *Science* **280**, 590–592 (1998).
- Trujillo, K. M., Yuan, S., Lee, E. Y. & Sung, P. Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. *J. Biol. Chem.* **273**, 21447–21450 (1998).
- Paull, T. T. & Gellert, M. Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes Dev.* **13**, 1276–1288 (1999).
- Maser, R. S., Monsen, K. J., Nelms, B. E. & Petrini, J. H. hMre11 and hRad50 nuclear foci are induced during the normal cellular response to DNA double-strand breaks. *Mol. Cell. Biol.* **17**, 6087–6096 (1997).
- Maser, R. S., Mressan, D. A. & Petrini, J. H. *The Mre11-Rad50 Complex: Diverse Functions in the Cellular DNA Damage Response* (Humana, Totowa, 1999).

- Khosravi, R. *et al.* Rapid ATM-dependent phosphorylation of MDM2 precedes p53 accumulation in response to DNA damage. *Proc. Natl Acad. Sci. USA* **96**, 14973–14977 (1999).
- Matsuoka, S., Huang, M. & Elledge, S. J. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science* **282**, 1893–1897 (1998).
- Brown, A. L. *et al.* A human Cds1-related kinase that functions downstream of ATM protein in the cellular response to DNA damage. *Proc. Natl Acad. Sci. USA* **96**, 3745–3750 (1999).
- Cortez, D., Wang, Y., Qin, J. & Elledge, S. J. Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. *Science* **286**, 1162–1166 (1999).
- Zhong, Q. *et al.* Association of BRCA1 with the hRad50-hMre11-p95/nbs1 complex and the DNA damage response. *Science* **285**, 747–750 (1999).
- Ziv, Y. *et al.* Recombinant ATM protein complements the cellular A-T phenotype. *Oncogene* **15**, 159–167 (1997).
- Morgan, S. E., Lovly, C., Pandita, T. K., Shiloh, Y. & Kastan, M. B. Fragments of ATM which have dominant-negative or complementing activity. *Mol. Cell. Biol.* **17**, 2020–2029 (1997).

Acknowledgements

We thank T. M. Gilmer and K. Kilpatrick (Glaxo-Wellcome) for the mouse ATM monoclonal antibodies; Y. Shiloh for AT complemented cell lines; C. Bakkenist, K. McLure and C. Canman for helpful comments on the manuscript; and D. Woods and A. Justman for technical assistance. This work was supported by grants from the NIH (M.B.K., the St. Jude Cancer Centre Core Grant, J.H.J.P. and J.L.). M.B.K. was also supported by Glaxo-Wellcome and the American Lebanese Syrian Associated Charities (ALSAC) of the St. Jude Children's Research Hospital, and J.H.J.P. by the Milwaukee Foundation.

Correspondence and requests for materials should be addressed to M.B.K (e-mail: Michael.Kastan@stjude.org).

Integrin LFA-1 interacts with the transcriptional co-activator JAB1 to modulate AP-1 activity

Elisabetta Bianchi*, Simona Denti*, Alessandra Granata*, Giovanna Bossi†‡, Jens Geginat*, Antonello Villa*§, Lars Rogge|| & Ruggero Pardi*¶

* Scientific Institute San Raffaele-DIBIT, I-20132 Milano, Italy

† Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK

§ MIA, Facoltà di Medicina, Univ. di Milano Bicocca, I-20052 Monza, Italy

|| Roche Milano Ricerche, I-20132 Milano, Italy

¶ University Vita-Salute San Raffaele, I-20132 Milano, Italy

Integrin adhesion receptors transduce signals that control complex cell functions which require the regulation of gene expression, such as proliferation, differentiation and survival¹. Their intracellular domain has no catalytic function, indicating that interaction with other transducing molecules is crucial for integrin-mediated signalling. Here we have identified a protein that interacts with the cytoplasmic domain of the β2 subunit of the αL/β2 integrin LFA-1. This protein is JAB1 (Jun activation domain-binding protein 1), a coactivator of the c-Jun transcription factor². We found that JAB1 is present both in the nucleus and in the cytoplasm of cells and that a fraction of JAB1 colocalizes with LFA-1 at the cell membrane. LFA-1 engagement is followed by an increase of the nuclear pool of JAB1, paralleled by enhanced binding of c-Jun-containing AP-1 complexes to their DNA consensus site and increased transactivation of an AP-1-dependent promoter. We suggest that signalling through the LFA-1 integrin may affect c-Jun-driven transcription by regulating JAB1 nuclear localization. This represents a new pathway for integrin-dependent modulation of gene expression.

‡ Present address: Department of Experimental Oncology, European Institute of Oncology, via Ripamonti 435, Milan 20141, Italy.

Integrin cytoplasmic domains are crucial in the establishment of organized complexes with the cytoskeleton³, which contain many structural and signalling proteins¹ and are important for integrin-mediated signal transduction. To identify molecules directly involved in transducing adhesion-generated signals, we used the intracellular domain of the LFA-1 $\beta 2$ chain ($\beta 2$ cyt) fused to the GAL4 DNA-binding domain (DBD) as a bait in the yeast two-hybrid system to screen a human lymphocyte complementary DNA library⁴. We identified two independent, overlapping cDNAs that specifically potentiated activation of a GAL4-dependent reporter in the presence of the $\beta 2$ cyt-DBD fusion, but not with other control DBD fusions, including the cytoplasmic domain of LFA-1 αL chain (αL cyt, data not shown). The cDNAs encode JAB1, a transcriptional coactivator for the Jun family of AP-1 transcription factors².

In vitro transcribed and translated JAB1 bound to a glutathione-S-transferase (GST)- $\beta 2$ cyt fusion protein, but not to GST alone nor to other unrelated GST fusions, including GST- αL cyt, GST-p53 and GST-*trk* B (Fig. 1a and data not shown), confirming the above results. JAB1 also bound to a GST-c-Jun activation domain (1-79) fusion, predictably as JAB1 binds to the activation domain of specific Jun proteins² (Fig. 1a). The same GST- $\beta 2$ cyt protein, but not GST- αL cyt nor GST alone, specifically co-precipitated the endogenous JAB1 protein from cytoplasmic extracts (Fig. 1b).

We then asked whether the association between $\beta 2$ cyt and JAB1 resulted in colocalization of the two proteins *in vivo*. We transiently expressed the integrin LFA-1 and epitope-tagged JAB1 (Flag-JAB1) in mammalian cells (COS7), and visualized the proteins by double immunofluorescence staining. When cells were plated on a substrate that does not ligate LFA-1 (fibronectin), the ectopically expressed integrin concentrated at peripheral areas of the cell membrane, resembling ruffles (Fig. 2a, top and middle panels, red), where it associates with cytoskeletal elements⁵. A fraction of JAB1 was also enriched in the same peripheral membrane regions (Fig. 2a, top and middle panels, green); the rest was distributed heterogeneously in the nucleus and in the cytoplasm. In cells not expressing LFA-1, JAB1 maintained a diffuse distribution (Fig. 2a, top and bottom panels). In normal T lymphocytes (Fig. 2c), a

fraction of the endogenous JAB1 colocalizes with LFA-1, but not with another cell surface molecule (LFA-3, Fig. 2c).

As receptor ligation and clustering are required to initiate integrin-dependent signalling events, we tested whether engagement of LFA-1 by specific monoclonal antibody (mAb) could affect intracellular localization of JAB1. We stained COS7 cells transfected with JAB1 and LFA-1 cDNAs with an anti- $\beta 2$ mAb and allowed them to spread at 37 °C on immobilized anti- αL mAb, to mimic ligand-induced clustering of the receptor. Thirty minutes after plating we observed clustering of LFA-1 (Fig. 2b, arrowheads), loss of membrane-associated JAB1 and redistribution of part of JAB1 to the integrin clusters. The rest of JAB1 was present mainly in the cytoplasm, with only a small fraction of cells showing nuclear localization (range 5–18% in three separate experiments). Sixty minutes after stimulation there was a clear increase in the fraction of positive nuclei (range 22–38% in three independent experiments, Fig. 2b). In a control experiment, where cells were transfected with the ICAM1 receptor cDNA and adhered to a plate coated with an anti-ICAM1, JAB1 was diffused throughout the cell without evident nuclear enrichment (Fig. 2b, bottom panels) and the fraction of cells with positive nuclei remained constant over time (not shown).

We confirmed these morphological findings by western blot analysis of JAB1 content in cytoplasmic and nuclear fractions of transfected COS7 cells. JAB1 increased with time in the nuclear extracts when COS7 cells were stimulated by crosslinking LFA-1 (3.4 ± 1.2 -fold increase at 60 min, mean and s.d. of three experiments), but not when ICAM1 was crosslinked (Fig. 3a).

We next investigated whether there is an increase in JAB1 nuclear pool upon engagement of endogenous LFA-1 in normal primary T cells. Crosslinking of LFA-1 induced a rapid increase in endogenous JAB1 in the nuclear fraction, as compared with unstimulated lymphocytes (2.5 ± 0.6 -fold at 10 min, mean and s.d. of six independent experiments, Fig. 3b, c). Ligation of other surface molecules (CD4, CD8 and LFA-3) did not influence JAB1 subcellular distribution (Fig. 3c). Similar results, although with slightly different kinetics, were obtained by binding to a physiological LFA-1 ligand (immobilized recombinant ICAM1) (Fig. 3d). Increased levels of nuclear JAB1 were also detected by indirect immunofluorescence in a fraction of T cells and U937 cells adhered to immobilized anti-LFA-1 mAb (Fig. 3e).

Overexpression of JAB1 potentiates transactivation of the AP-1-dependent collagenase promoter². The mechanism proposed for JAB1 induced coactivation is the stabilization of c-Jun complexes with AP-1 sites². We therefore asked whether the observed increase in the JAB1 nuclear pool which follows LFA-1 stimulation was paralleled by variations in the binding of nuclear AP-1 complexes to an AP-1 site. We prepared nuclear extracts from COS7 cells transfected with LFA-1 and JAB1 and stimulated with anti-LFA-1 mAb and used them in electrophoretic mobility-shift assays (EMSA) to test their ability to bind to a labelled AP-1 probe *in vitro*. Crosslinking of LFA-1 for 60 min resulted in an increase in the amount of bound AP-1 probe (1.5 to 3-fold, in four independent experiments) as compared with AP-1 binding at 0 and 30 min after stimulation (Fig. 4a, lanes 1–3). Crosslinking of LFA-1 in COS7 cells expressing only endogenous levels of JAB1 also resulted in increased binding of the AP-1 probe (Fig. 4a, lanes 4–6), but the overall amount of bound AP-1 complexes was lower. To test for equal protein loading in all lanes, we analysed nuclear extracts by western blotting of Retinoblastoma (Rb) protein content (Fig. 4a, bottom panel). The protein-DNA complex was completely supershifted by pre-incubation with anti-c-Jun or anti-c-Fos mAb, indicating that they consist of c-Jun/c-Fos heterodimers (Fig. 4a, lanes 7 and 8).

We next investigated the effects of JAB1 overexpression and LFA-1 stimulation on AP-1-dependent transcriptional activity. We transfected COS7 cells with an AP-1-driven luciferase reporter gene⁶ in the presence of c-Jun, JAB1 and either LFA-1 or ICAM1. Cells were

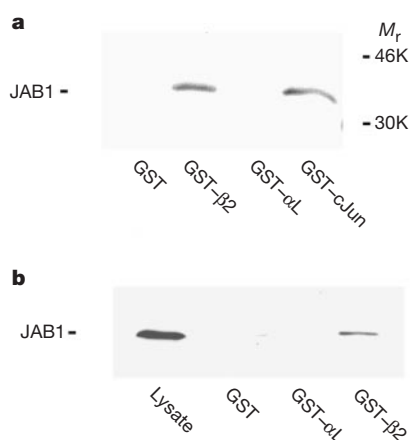


Figure 1 JAB1 binds to the cytoplasmic domain of the LFA-1 $\beta 2$ chain. **a**, *In vitro* transcribed and translated JAB1 was incubated with immobilized glutathione-S-transferase (GST) or with the indicated GST-fusion proteins. Bound JAB1 was resolved on SDS-PAGE and visualized by autoradiography. GST- $\beta 2$: cytoplasmic domain of LFA-1 $\beta 2$ chain fused to GST; GST- αL : GST fusion of LFA-1 αL chain cytoplasmic domain; GST-c-Jun: GST fusion of c-Jun activation domain (1–79). **b**, Cytoplasmic extracts were incubated with immobilized GST, GST- αL or GST- $\beta 2$ fusion proteins. Bound proteins were eluted with excess glutathione and analysed by immunoblotting with anti-JAB1 monoclonal antibodies.

stimulated by adhesion to either anti-LFA-1 or anti-ICAM1 mAb and assayed for luciferase activity. Overexpression of JAB1 caused an increase in the relative activation levels of the AP-1 reporter (from 1 to 1.9 ± 0.8 , mean $\pm$ s.d. of four separate experiments for ICAM1-stimulated cells; Fig. 4b, white bars). The increase was much more substantial, however, when JAB1 was expressed in cells stimulated by LFA-1 crosslinking (from 2.9 ± 0.6 to 14 ± 4.7 ; Fig. 4b, grey bars), indicating a synergy between LFA-1-mediated signalling and high levels of JAB1. The same synergy was observed in Jurkat T cells, in the presence of overexpressed JAB1 and after LFA-1 stimulation (Fig. 4c). LFA-1 crosslinking in the absence

of exogenous JAB1 induced a threefold potentiation of AP-1 activity compared with ICAM1 stimulation (Fig. 4b). This may be due to a redistribution of endogenous JAB1, but also to integrin-induced activation of Jun kinase (JNK) and phosphorylation of c-Jun⁷⁻⁹. In our system, JAB1 was also capable of potentiating transactivation by a phosphorylation-defective c-Jun mutant, although less efficiently than in the case of wild-type c-Jun (Fig. 4d). This indicates that phosphorylation of c-Jun is synergistic, but not indispensable for JAB1 activity.

Overexpression of two different JAB1 antisense constructs resulted in a clear reduction of the endogenous JAB1 levels

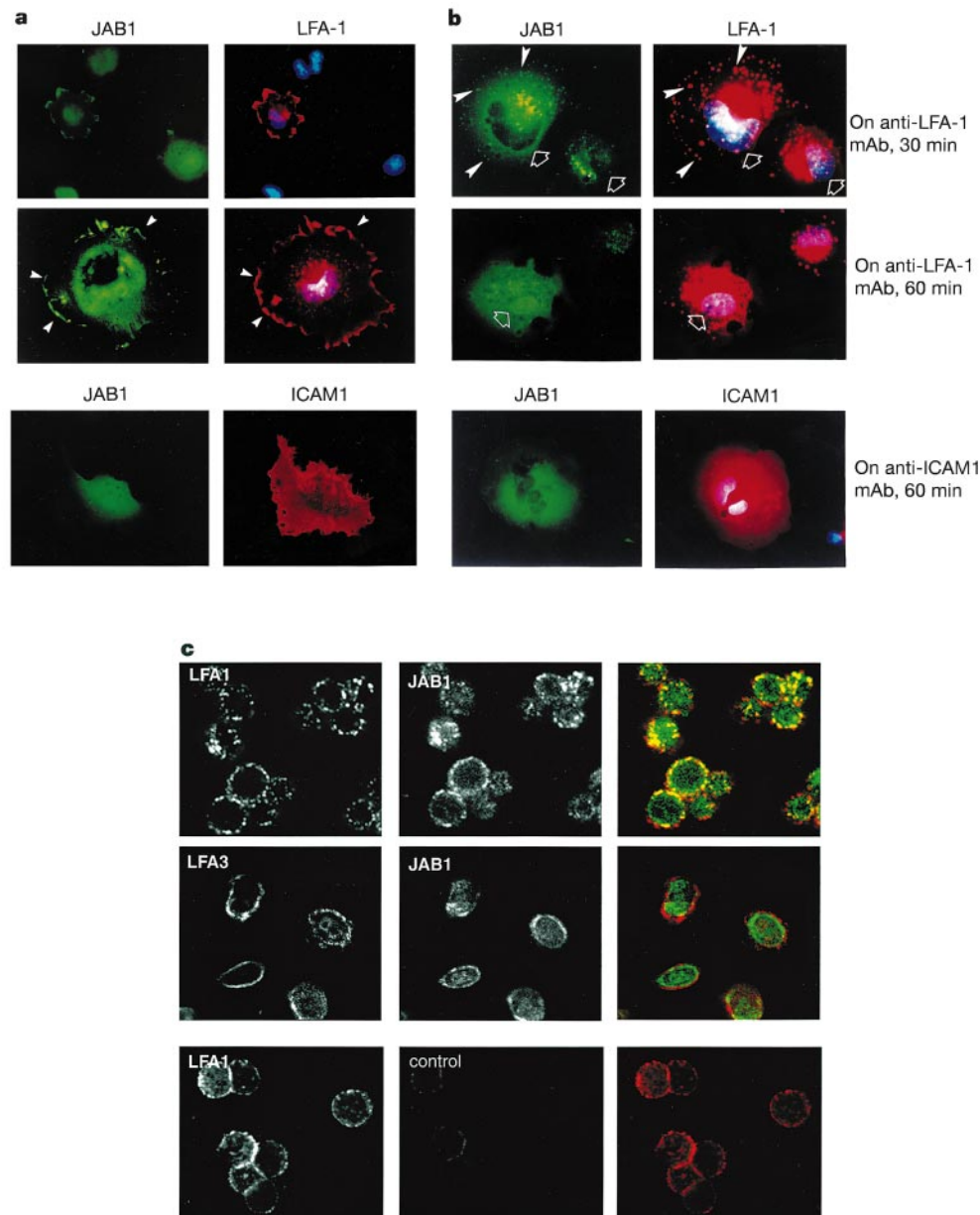


Figure 2 JAB1 colocalizes with LFA-1 in COS7 cells and is enriched in the nucleus upon LFA-1 aggregation. **a**, COS7 cells were co-transfected with JAB1 and with LFA-1 (top and middle) or ICAM1 cDNAs (bottom) and plated on fibronectin. LFA-1, ICAM1 (red) and JAB1 (green) were detected by indirect immunofluorescence. Arrowheads indicate membrane ruffle-like areas, where both LFA-1 and JAB1 concentrate. **b**, Cells stained for LFA-1 or ICAM1 (red) were adhered for 30 or 60 min to anti- α L (top panels) or anti-ICAM1 mAb (bottom panels, 60 min). JAB1 was detected as above (green). Outlined arrows indicate

nuclei stained by Hoechst 33342 (blue), arrowheads indicate clusters of JAB1 colocalizing with clustered LFA-1 integrin. **c**, A fraction of the endogenous JAB1 colocalizes with LFA-1 in T lymphocytes. Top two rows: cells were labelled with anti-LFA-1 or anti-LFA-3 mAb (left) and with anti-JAB1 mAb (middle). Samples were analysed by confocal microscopy. Right, merge of JAB1 (green) and LFA-1 or LFA-3 (red) images. Bottom row: cells were stained with anti-LFA-1 mAb (left) and an irrelevant mAb (middle); right, merge of left and middle panels (red: LFA-1, green: irrelevant mAb).

(Fig. 4f) and partially inhibited the activation of the AP-1 reporter gene induced by LFA-1 crosslinking (Fig. 4e). This indicates that endogenous JAB1 is physiologically involved in the regulation of AP-1 transcriptional activity.

Integrins cooperate with growth factors in the induction of AP-1 transactivating function¹⁰. Our data support the hypothesis that one of the pathways by which stimuli delivered through LFA-1 modulate AP-1 transcription is the redistribution of JAB1 from an extra-nuclear to a nuclear compartment. Although the mechanism responsible for the relocalization of JAB1 upon integrin clustering remains to be identified, it is interesting that a direct translocation from the cell membrane to the nucleus in response to extracellular stimuli has been described for a number of components or regulators of the transcriptional machinery, such as STAT¹¹ or SMAD transcription factors¹², β -catenin¹³ and c-Abl¹⁴.

JAB1 has been isolated as a component of a highly conserved

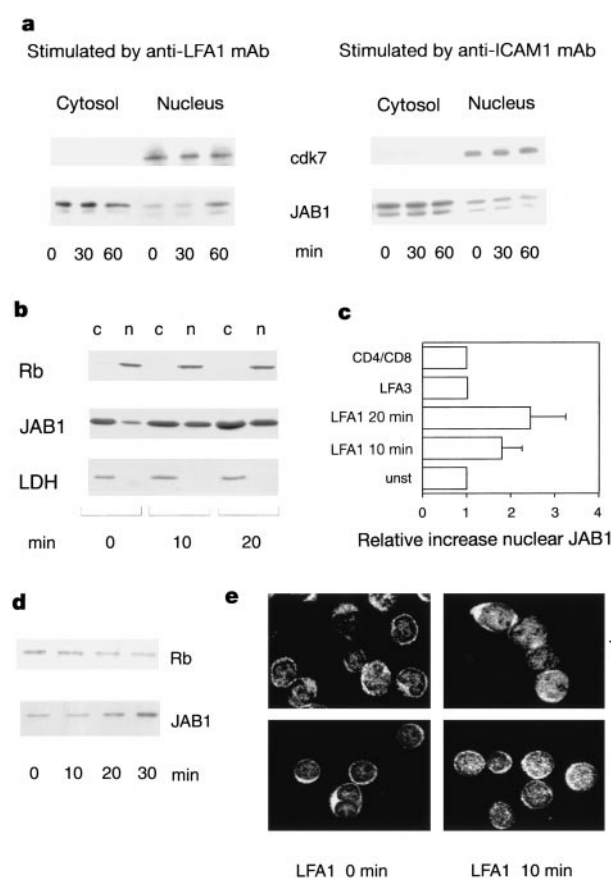


Figure 3 JAB1 is enriched in the nuclear fraction after LFA-1 stimulation. **a**, COS7 cells co-transfected with Flag-JAB1 and either LFA-1 or ICAM1 were stimulated by adhesion to anti-LFA-1 or to anti-ICAM1 mAb, respectively. Cytosolic and nuclear extracts were analysed by western blotting for nuclear cyclin-dependent kinase 7 (cdk7) and JAB1 content (the slower-migrating band represents Flag-JAB1, the faster band endogenous JAB1). **b**, Normal T lymphocytes were stimulated by adhesion to anti-LFA-1 mAb. Cytosolic (c) and nuclear (n) extracts were analysed by western blotting for JAB1 content and for nuclear (Rb: retinoblastoma protein) and cytoplasmic (LDH: lactic dehydrogenase) markers. **c**, Densitometric quantification of nuclear JAB1 levels in T lymphocytes stimulated by adhesion to anti-LFA-1 mAb or control (anti-CD4 plus anti-CD8) and anti-LFA-3 mAb (20 min), relative to the unstimulated control (unst.). Bars represent mean and s.d. of six experiments. **d**, Western blot analysis of JAB1 and Rb content in the nuclear fractions of normal T cells, stimulated by adhesion to immobilized recombinant ICAM1. **e**, T lymphocytes (top) or U937 cells (bottom) were adhered to immobilized anti-LFA-1 mAb and incubated at 37 °C for the indicated times. JAB1 was visualized by indirect immunofluorescence and confocal microscopy.

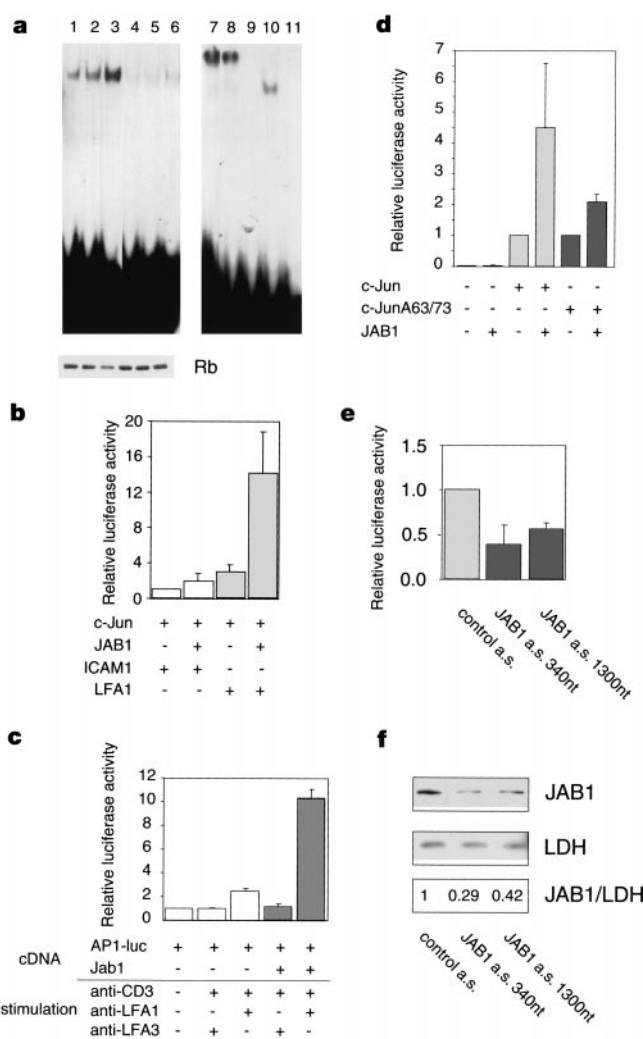


Figure 4 LFA-1 stimulation synergizes with JAB1 in inducing AP-1 transcriptional activity. **a**, COS7 cells were transfected with LFA-1 and JAB1 cDNAs (lanes 1–3 and 7–11) or with LFA-1 alone (lanes 4–6) and stimulated with anti-LFA-1 mAb. Nuclear extracts were collected at 0 (lanes 1, 4), 30 (lanes 2, 5) and 60 min (lanes 3, 6, 7, 8–11) and tested for their ability to bind to a ³²P-labelled AP-1 oligonucleotide. Control samples were pre-incubated with an anti-c-Jun mAb (lane 7), an anti-c-Fos mAb (lane 8), a 100-fold excess of unlabelled AP-1 oligonucleotide (lane 9) or a 100-fold excess of an unrelated oligonucleotide (lane 10). Cytoplasmic extracts did not bind the AP-1 oligo (lane 11). As a loading control, Rb content in the nuclear extracts used in the EMSA was analysed by immunoblot (bottom). **b**, COS7 cells were transfected with an AP-1-driven luciferase reporter gene and with the indicated cDNAs and stimulated by adhesion to immobilized anti-ICAM1 (white bars) or anti-LFA-1 (grey bars) mAb. Relative luciferase activity in the whole cell extracts is shown (mean and s.d. of four independent experiments). **c**, Jurkat T cells were transfected with (grey bars) or without (white bars) JAB1 cDNA and with an AP-1-driven luciferase reporter gene. Samples were stimulated with anti-LFA-1 or anti-LFA-3 mAb. Normalized luciferase activity is represented relative to the unstimulated control. **d**, Contribution of c-Jun phosphorylation to LFA-1-induced AP-1 activity. COS7 cells were transfected with the AP-1 luciferase reporter gene and the indicated cDNAs (white bars: no c-Jun; grey bars: c-Jun; black bars: c-JunA63/73) and stimulated by adhesion to anti-LFA-1 mAb. c-JunA63/73 is a phosphorylation defective c-Jun mutant. Bars represent mean and s.d. of three experiments. **e**, COS7 cells were transfected with LFA-1, c-Jun, AP-1-luciferase reporter and one of the indicated antisense (a.s.) constructs. Cells were stimulated by adhesion to anti-LFA-1 mAb. Bars represent mean and s.d. of luciferase activity from six (for 1,300-nucleotide (nt) a.s.) or three (340-nucleotide a.s.) independent transfections, relative to control antisense samples. **f**, Western blot analysis of JAB1 in the antisense transfected samples. Numbers express JAB1 to LDH ratios, as obtained by densitometry scanning.

multimolecular complex (the 'signalosome' or COP9 complex) with kinase activity towards different transcriptional regulators^{15,16}. Notably, plant COP9 is a photomorphogenesis regulator which also has been described to shuttle between the cytoplasm and the nucleus in response to extracellular stimuli¹⁷. Ectopically expressed JAB1 promotes nucleocytoplasmic transport of the cyclin-dependent kinase inhibitor p27 (ref. 18). Whether cellular redistribution of JAB1 occurs within a multiprotein complex remains to be determined, but it is intriguing to speculate that the signalosome could represent a crossroad, where signals originating from the extracellular environment are coordinated with transcriptional activation and with the regulation of cell-cycle progression. □

Methods

Two-hybrid system

β 2 and α L cytoplasmic domains were fused to GAL4-DBD (1–147) in the pAS1 vector⁴. Screening of an EBV-transformed B-cell library was conducted as described⁴ in the presence of 25 mM 3-aminotriazole. Two independent clones were isolated that interacted with GAL4-DBD- β 2cyt, but not with unrelated baits (α Lcyt, p53, largeT antigen, the catalytic subunit and the p49 primase subunit of DNA polymerase- α): one encoding amino acids 44–334 and the other the full-length JAB1 protein.

Protein-protein interaction assays

β 2cyt, α Lcyt and p53 were subcloned into pGEX4T1 (Pharmacia). GST fusions were purified according to the manufacturer's protocol (Pharmacia). Immobilized GST fusions were incubated with 5 μ l ³⁵S-labelled, *in vitro* transcribed and translated JAB1 (Promega), washed with 400 volumes of binding buffer (20 mM HEPES pH 7.4, 200 mM NaCl, 0.25 mM MgCl₂, 0.5 mM DTT, 1% NP40), eluted with 25 mM glutathione, resolved on SDS-PAGE and visualized by autoradiography. For 'pull-down' assays, cytoplasmic fractions were extracted with 0.025% digitonin, incubated 60 min at 4 °C with GST fusions, immobilized on Sepharose beads and washed twice with 50 volumes of the same buffer. Bound proteins were visualized by western blotting with anti-JAB1 mAb. One-fourth of input lysate was loaded as control.

In vivo expression and immunofluorescence studies

Anti-JAB1 mAb (SP682 and SP602, IgG2b) were raised against bacterially expressed JAB1. pFLAG-CMV2-JAB1, pCDM8- α L or β 2, pRC-ICAM1 were transfected by calcium phosphate in COS7 cells. For colocalization studies, transfected cells were plated on fibronectin. LFA-1 was detected by mAb TS1.18 (anti- β 2, IgG1) and TS1.22 (anti- α L, IgG1), followed by Rhodamine-anti-IgG1 antiserum; JAB1 by mAb SP682 and SP602, followed by Fluorescein-anti-IgG2b. These conditions detected only overexpressed JAB1. In experiments where LFA-1 was engaged, transfected cells were stained with biotinylated TS1.18 mAb and Avidin-Cyanin3, adhered to TS1.22 mAb-coated plates and incubated at 37 °C for 30 and 60 min. ICAM1-transfected cells were adhered to anti-ICAM1 mAb (14D2D12, IgG1). ICAM1 was stained by incubation with biotinylated 1G12 mAb and Avidin-Cy3. After fixation and permeabilization JAB1 was detected as above. To detect endogenous JAB1 protein, purified human T lymphocytes were stained with TS1.18 or with TS2.9 (anti-LFA-3) mAb, followed by TRITC-anti-IgG1. After fixation, samples were stained with either SP602 or an irrelevant mAb, followed by biotinylated anti-IgG2b and Avidin-Cyanin2. Samples were analysed with an MRC-1024 confocal microscope (BioRad). To detect nuclear enrichment of JAB1, U937 cells and cells from a CD4+ T-cell line were adhered at 4 °C to anti-LFA-1-coated coverslips, then shifted at 37 °C for 10 min. Samples were fixed, permeabilized and stained for JAB1 as above.

Cell fractionation and western blot analysis

T cells were stimulated in suspension for 90 min with anti-CD3 mAb, then plated on immobilized anti-LFA-1 mAb or control mAb. COS7 cells were transfected with JAB1 and LFA-1 or ICAM1 cDNAs and stimulated by adhesion to antibody-coated plates. For stimulation on ICAM1, CD3-stimulated T cells were adhered to plates coated with recombinant α 1-ICAM1 as described¹⁹. To obtain cytoplasmic fractions, cells were treated for 7 min on ice with digitonin containing buffer (0.005% for lymphocytes and 0.0025% for COS7 cells). Nuclear fractions were then extracted with high salt buffer. Fraction purity was tested by western blotting, using as a cytoplasmic marker lactate dehydrogenase (LDH), and as a nuclear marker the Rb protein in T cells and c-Jun or cyclin-dependent kinase 7 (MO15, Zymed Laboratories)²⁰ in COS7 cells. JAB1 densitometry readings were normalized to the densitometry values of the nuclear marker in the same sample.

Electrophoretic mobility-shift assays

Nuclear extracts were obtained from transfected COS7 cells stimulated by adhesion to anti-LFA-1 antibodies, as described above. Five micrograms of proteins were incubated with a ³²P-labelled double-stranded oligonucleotide AP-1 probe as described². Protein-DNA complexes were separated on a 5% native polyacrylamide gel and visualized by autoradiography. Complexes were supershifted by pre-incubation (20 min at 4 °C) with the following mAb: c-Jun KM-1 (Santa Cruz Biotechnology) and c-Fos K-25-G (Santa Cruz Biotechnology).

Reporter gene assays

JAB1 cDNA was cloned into the pHOOK3 (Invitrogen) vector. COS7 cells were transfected in triplicate samples with an AP-1-driven luciferase reporter gene⁶ and with the indicated cDNAs, using the Fugene 6 reagent (Boehringer Mannheim). Phosphorylation-defective c-JunA63/73 was obtained from M. Karin²¹. Forty-eight hours after transfection, cells were adhered to plates coated with mAb against LFA-1 or ICAM1 and non-transfected cells removed by extensive washing. Adhered cells were allowed to spread for 6 h at 37 °C. Luciferase activity was analysed in cell lysates (Promega) and normalized to protein concentration. Jurkat T cells were transfected by a DEAE-dextran protocol. Cells were stimulated in suspension with anti-CD3 mAb, divided in two and plated on immobilized anti-LFA-1 or anti-LFA-3 mAb. Luciferase activity was normalized to β -galactosidase activity.

JAB1 full-length (1,300 nucleotides) and 1–340 nucleotide antisense fragments were cloned into pHook3. As a control, a 327-bp antisense cDNA from the cytoplasmic domain of the IL12 receptor β 2 subunit, not expressed in COS7 cells, was cloned in the same vector. Another unrelated control antisense construct (pcDNA-1.3 galactosyl transferase, 1,269 bp) gave similar results. COS7 cells were transfected in triplicates with cDNAs for the AP-1 luciferase reporter, c-Jun, LFA1 and the indicated antisense constructs. Twenty-four hours after transfection, cells were stimulated as above and luciferase activity and protein concentration were determined for each sample. JAB1 protein levels after antisense transfection were analysed by western blotting and compared with LDH levels in the same sample.

Received 2 December 1999; accepted 14 February 2000.

- Howe, A., Aplin, A. E., Alahari, S. K. & Juliano, R. L. Integrin signaling and cell growth control. *Curr. Opin. Cell Biol.* **10**, 220–231 (1998).
- Claret, F. X., Hibi, M., Dhut, S., Toda, T. & Karin, M. A new group of conserved coactivators that increase the specificity of AP-1 transcription factors. *Nature* **383**, 453–457 (1996).
- Burridge, K. & Chrzanowska-Wodnicka, M. Focal adhesion, contractility and signaling. *Annu. Rev. Cell. Dev. Biol.* **12**, 463–519 (1996).
- Durfee, T. et al. The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes Dev.* **7**, 555–569 (1993).
- Pardi, R., Bossi, G., Inverardi, L., Rovida, E. & Bender, J. R. Conserved regions in the cytoplasmic domains of the leukocyte integrin α L β 2 are involved in endoplasmic reticulum retention, dimerization, and cytoskeletal association. *J. Immunol.* **155**, 1252–1263 (1995).
- Rincon, M. & Flavell, R. A. AP-1 transcriptional activity requires both T-cell receptor-mediated and co-stimulatory signals in primary T lymphocytes. *EMBO J.* **13**, 4370–4381 (1994).
- Geginat, J., Bossi, G., Bender, J. R. & Pardi, R. Anchorage-dependence of mitogen-induced G1 to S transition in primary T lymphocytes. *J. Immunol.* **162**, 5085–5093 (1999).
- Miyamoto, S. et al. Integrin function: molecular hierarchies of cytoskeletal and signaling molecules. *J. Cell Biol.* **131**, 791–805 (1995).
- Mainiero, F. et al. The coupling of α 6 β 4 integrin to Ras-MAP kinase pathways mediated by Shc controls keratinocyte proliferation. *EMBO J.* **16**, 2365–2375 (1997).
- Tremble, P., Damsky, C. D. & Werb, Z. Components of the nuclear signaling cascade that regulate collagenase gene expression in response to integrin-derived signals. *J. Cell Biol.* **129**, 1707–1720 (1995).
- Leonard, W. J. & O'Shea, J. J. JAKS and STATS: biological implications. *Annu. Rev. Immunol.* **16**, 293–322 (1998).
- Kretschmar, M. & Massague, J. SMADs: mediators and regulators of TGF- β signaling. *Curr. Opin. Genet. Dev.* **8**, 103–111 (1998).
- Behrens, J. et al. Functional interaction of β -catenin with the transcription factor LEF-1. *Nature* **382**, 638–642 (1996).
- Lewis, J. M., Baskaran, R., Taagepera, S., Schwartz, M. A. & Wang, J. Y. Integrin regulation of c-Abl tyrosine kinase activity and cytoplasmic-nuclear transport. *Proc. Natl Acad. Sci. USA* **93**, 15174–15179 (1996).
- Seeger, M. et al. A novel protein complex involved in signal transduction possessing similarities to 26S proteasome subunits. *FASEB J.* **12**, 469–478 (1998).
- Wei, N. et al. The COP9 complex is conserved between plants and mammals and is related to the 26S proteasome regulatory complex. *Curr. Biol.* **8**, 919–922 (1998).
- Chamowitz, D. A. et al. The COP9 complex, a novel multisubunit nuclear regulator involved in light control of a plant developmental switch. *Cell* **86**, 115–121 (1996).
- Tomoda, K. et al. Degradation of the cyclin-dependent-kinase inhibitor p27^{Kip1} is instigated by Jab1. *Nature* **398**, 160–165 (1999).
- Fabbri, M. et al. A tyrosine-based sorting signal in the β 2 integrin cytoplasmic domain mediates its recycling to the plasma membrane and is required for ligand-supported migration. *EMBO J.* **18**, 4915–4925 (1999).
- Tassan, J. P., Schultz, S. J., Bartek, J. & Nigg, E. A. Cell cycle analysis of the activity, subcellular localization, and subunit composition of human CAK. *J. Cell Biol.* **127**, 467–478 (1994).
- Smeal, T., Binetruy, B., Mercola, D. A., Birrer, M. & Karin, M. Oncogenic and transcriptional cooperation with Ha-Ras requires phosphorylation of c-Jun on serines 63 and 73. *Nature* **354**, 494–496 (1991).

Acknowledgements

We thank S. Putignano for generating anti-JAB1 MoAb, S. Biffo for GST-TrkB, M. Karin for the GST-c-Jun and the c-JunA63/73 constructs, A. Cassetti for help with confocal microscopy, S. Trinca for administrative assistance and V. Zimarino for critically reviewing the manuscript. This work was supported by grants from AIRC, Telethon and MURST (to R.P.).

Correspondence and requests for materials should be addressed to E.B. (e-mail: bianchi.elisabetta@hsr.it).

XCDT1 is required for the assembly of pre-replicative complexes in *Xenopus laevis*

Domenico Maiorano*, Jacques Moreau† & Marcel Méchali*

* Institute of Human Genetics, CNRS, Genome Dynamics and Development, 141, rue de la Cardonille, 34396 Montpellier, Cedex 5, France

† Institut Jacques Monod, Université Paris VII, 2 Place Jussieu, Paris Cedex 05, France

In eukaryotic cells, chromosomal DNA replication begins with the formation of pre-replication complexes at replication origins. Formation and maintenance of pre-replication complexes is dependent upon CDC6 (ref. 1), a protein which allows assembly of MCM2–7 proteins^{2–4}, which are putative replicative helicases⁵. The functional assembly of MCM proteins into chromatin corresponds to replication licensing. Removal of these proteins from chromatin in S phase is crucial in origins firing regulation⁶. We have identified a protein that is required for the assembly of pre-replication complexes, in a screen for maternally expressed genes in *Xenopus*. This factor (XCDT1) is a relative of fission yeast cdt1, a protein proposed to function in DNA replication⁷, and is the first to be identified in vertebrates. Here we show, using *Xenopus in vitro* systems, that XCDT1 is required for chromosomal DNA replication. XCDT1 associates with pre-replicative chromatin in a manner dependent on ORC protein and is removed from chromatin at the time of initiation of DNA synthesis. Immunodepletion and reconstitution experiments show that XCDT1 is required to load MCM2–7 proteins onto pre-replicative chromatin. These findings indicate that XCDT1 is an essential component of the system that regulates origins firing during S phase.

A *Xenopus* gene (*XCDT1*) encoding a protein with significant homology to the fission yeast cdt1 protein⁷ was isolated from a complementary DNA library made with maternally enriched RNAs, in a search for early expressed genes in *Xenopus* embryos⁸. Fission

yeast *cdt1*⁺ is an essential gene transcribed at the onset of S phase together with *cdc22*⁺ (encoding the large subunits of ribonucleotide reductase⁹) and *cdc18*⁺ (ref. 10), a *XCDT1* homologue⁴. XCDT1 (GenBank accession number AJ250122) is 24% identical and 40% similar to yeast cdt1, which is like the identity observed between yeast and *Xenopus* ORC proteins (20%; ref. 11), suggesting that the *Xenopus* cdt1-related protein may be a true homologue of yeast cdt1. No obvious homologies with proteins from budding yeast were found, whereas human and mouse expressed sequence tag displaying high homology with XCDT1 were identified (60% identity, data not shown), suggesting that cdt1 may be widely conserved in metazoans.

The predicted XCDT1 protein (relative molecular mass 69,800 (*M_r* 69.8K)) is larger than yeast cdt1 (*M_r* 54K). This difference is due to an amino-terminal extension of 176 amino acids (Fig. 1a), which also contains several potential phosphorylation sites ('P' in Fig. 1a) for cyclin-dependent kinases. Both XCDT1 and fission yeast cdt1 are basic proteins (isoelectric point (pI) ≈ 10) and contain two regions with a high probability of assuming a coiled-coil conformation, suggesting potential protein–protein interaction domains.

Polyclonal antibodies raised against recombinant XCDT1 protein recognize the *in vitro* translated protein (Fig. 1b) and a single polypeptide of *M_r* 75K in *Xenopus* egg extracts ('HSS' in Fig. 1b). XCDT1 has a slower mobility in extracts prepared from unfertilized *Xenopus* eggs ('CSF' in Fig. 1c), compared with interphasic extracts ('-OA' in Fig. 1c). This discrepancy is probably due to phosphorylation as, following release in interphase, the slow-migrating form of XCDT1 is stabilized by okadaic acid, a phosphatase inhibitor ('+OA' in Fig. 1c).

The function of fission yeast cdt1 is unknown, although its genetic characterization suggests that it could be required for

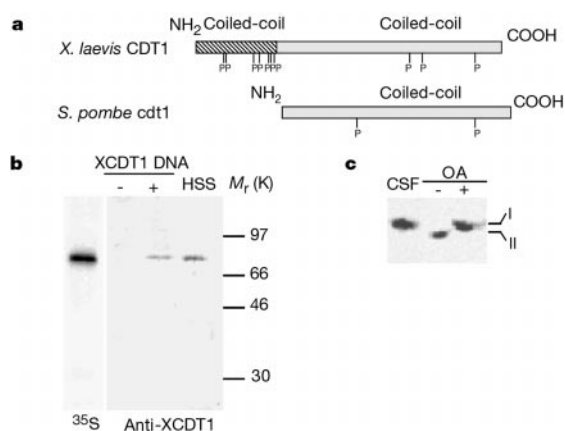


Figure 1 *Xenopus* CDT1 is homologous to fission yeast cdt1 protein. **a**, Diagram of predicted proteins. Putative phosphorylation sites by cyclin-dependent kinases (P) are indicated. **b**, Characterization of XCDT1-specific antibodies. Autoradiography of *in vitro* translated, ³⁵S-labelled XCDT1 (³⁵S), see Methods, after SDS–PAGE. One aliquot of the reaction without (–) or with (+) XCDT1 DNA was analysed by western blotting with XCDT1-specific antibody. HSS, *Xenopus* S-phase high speed egg extract. **c**, XCDT1 isoforms in metaphase-arrested *Xenopus* egg extracts. Unfertilized egg extract (CSF) was released into interphase in the absence (–) or presence (+) of 1 μM okadaic acid (OA). Protein samples were resolved by 10% SDS–PAGE and analysed by western blotting with XCDT1-specific antibody. I and II indicate two isoforms of XCDT1.

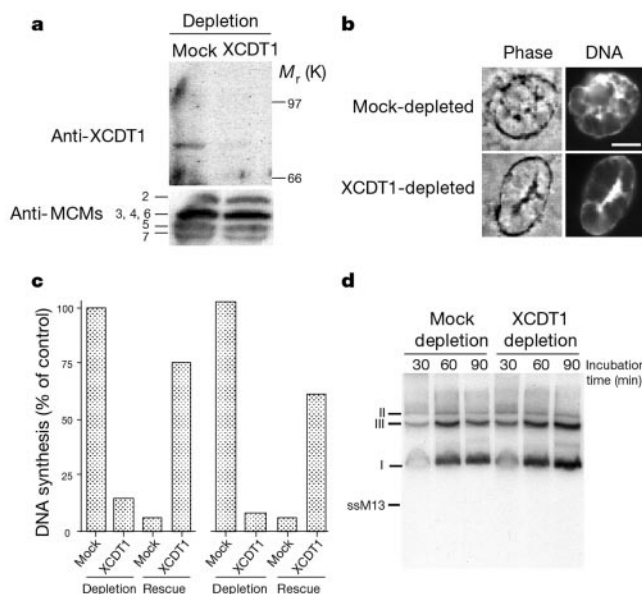


Figure 2 *Xenopus* CDT1 is required for chromosomal DNA replication. **a**, Depletion of XCDT1 from *Xenopus* egg extracts. Western blotting of egg supernatants depleted with either control (Mock) or XCDT1-specific (XCDT1) antibodies (90% of control). Numbers on the left-hand side of the lower panel indicate MCMs. **b**, Nuclear assembly in XCDT1-depleted extracts. Nuclei were observed by phase-contrast microscopy and stained with Hoechst to visualize DNA. Scale bar, 10 μm. **c**, DNA replication in XCDT1-depleted extracts. Left panel, replication of sperm chromatin (90-min incubation) in Mock- or XCDT1-depleted extracts and rescue with proteins eluted from mock or XCDT1-specific antibodies. Right panel, rescue with recombinant XCDT1. **d**, Time course of single-stranded M13 DNA replication in XCDT1-depleted extracts. Autoradiography after agarose gel electrophoresis of the reaction products. The mobility of unlabelled, single-stranded (ssM13), supercoiled (I), relaxed (II) and linear (III) forms of M13 are indicated.

DNA replication⁷. In *Xenopus*, S phase can be reproduced *in vitro* using extracts derived from activated *Xenopus* eggs. To determine whether XCDT1 is required for DNA replication in *Xenopus*, the protein was removed from egg extracts using specific antibodies (Fig. 2a), and the depleted extracts were challenged to support nuclear assembly and DNA replication *in vitro*. Nuclei assembled normally in XCDT1-depleted extracts (Fig. 2b); however, the replication of sperm chromatin was markedly reduced (Fig. 2c). The DNA synthesis defect of XCDT1-depleted extracts could be recovered by addition of proteins eluted from XCDT1-specific antibodies ('XCDT1' rescue in Fig. 2c, left panel) but not from control antibodies ('Mock' rescue in Fig. 2c, left panel), indicating that the depletion treatment did not irreversibly inactivate the competence of the extracts to support DNA synthesis. A similar result was obtained by the addition of a recombinant XCDT1 protein (Fig. 2c, right panel) showing that failure to activate S phase in XCDT1-depleted extracts is not caused by removal of associated proteins. XCDT1-depleted extracts efficiently replicated ssM13 DNA (Fig. 2d), as in untreated extracts¹², suggesting that in the absence of XCDT1 processes occurring at the replication fork such as priming and elongation of DNA chains are not affected. The

formation of supercoiled ssM13 DNA (form 'I' in Fig. 2d) indicates that replicated DNA is also efficiently assembled into chromatin, similarly to untreated extracts¹³. Taken together, these results indicate that the XCDT1 protein is likely to be required at the initiation step of DNA replication. This is consistent with the phenotype of *cdt1* mutants in fission yeast⁷ and suggests that XCDT1 may be a conserved vertebrate homologue of fission yeast *cdt1*.

The fission yeast *cdt1* protein has not been biochemically characterized and its subcellular location is unknown⁷. We determined whether XCDT1 would be enriched in the nucleus, as expected for a DNA replication protein. Upon incubation in egg extracts, *Xenopus* demembrated sperm nuclei quickly become competent to replicate (10–15 min, the licensing reaction¹⁴) and subsequently acquire a nuclear membrane (30–40 min) which allows initiation of DNA synthesis. Analysis of chromatin formed in interphasic egg extracts during early S phase shows that XCDT1, which is absent from isolated sperm chromatin ('Sc' in Fig. 3a), associates very rapidly with sperm chromatin (1–3 min), with kinetics similar to XMCM4 (ref. 15). Unlike XMCM4, accumulation of XCDT1 on chromatin at

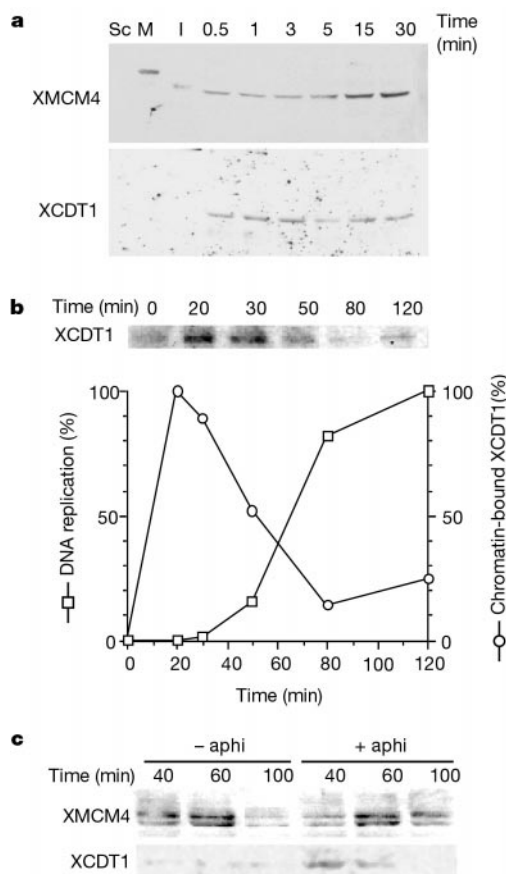


Figure 3 XCDT1 is a chromatin-associated protein. **a**, Binding of XCDT1 to sperm chromatin. *Xenopus* interphasic egg extracts (low speed supernatant) were incubated with demembrated sperm nuclei and chromatin was purified (see Methods) at the indicated time points. Bound proteins were separated by 7.5% SDS-PAGE and analysed by western blotting. Sc, sperm chromatin; M, mitotic egg extract; I, interphasic egg extract. **b**, XCDT1 is removed from chromatin upon activation of DNA synthesis. Chromatin obtained as in **a** was isolated at the indicated times and analysed by western blotting. DNA synthesis was monitored in parallel and recorded in the plot together with the quantification of the western blot signals of the upper panel. **c**, XCDT1 is transiently stabilized on chromatin by aphidicolin. Western blotting of chromatin isolated at the indicated time points in the absence (-aphi) or presence (+aphi) of aphidicolin (50 μ g ml⁻¹). XMCM4 migrates as a doublet (two phosphorylation isoforms).

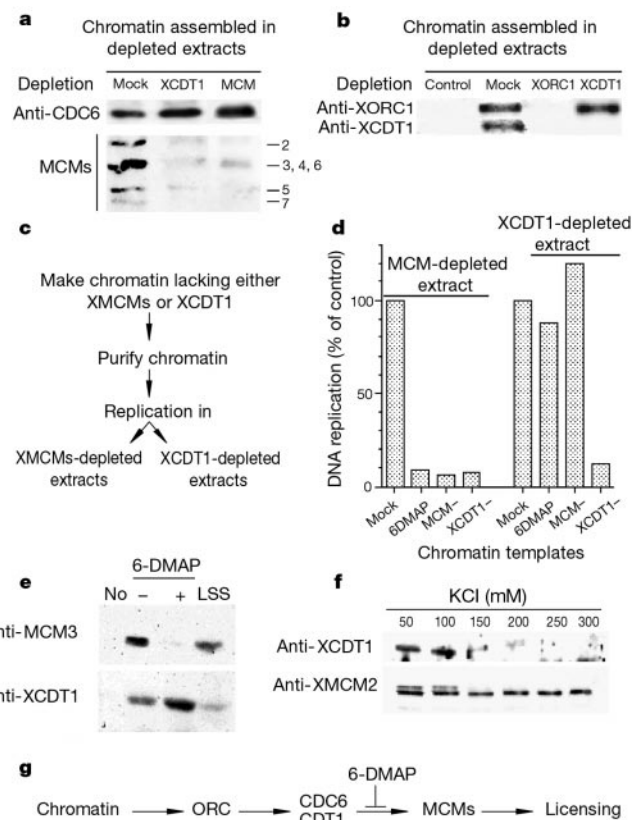


Figure 4 XCDT1 is required for the assembly of MCM proteins onto pre-replicative chromatin. **a**, **b**, Proteins bound to XCDT1-depleted pre-replicative chromatin. Western blotting of proteins eluted from chromatin assembled in mock-, XCDT1-, ORC1-, or MCM-depleted extracts. Numbers indicate MCM proteins. Control is insoluble fraction obtained by centrifugation of egg extracts without sperm nuclei added. Over 90% depletion of the corresponding proteins was achieved in these experiments. **c**, Diagram of the experimental protocol of **d**, **d**, Replication of chromatin templates in MCMs- and XCDT1-depleted extracts. Mock chromatin is chromatin made in mock-depleted extracts. Replication was measured after 90-min incubation at 23 °C. **e**, XCDT1 is bound to 6-DMAP-treated chromatin. Western blotting of proteins eluted from chromatin formed in egg extracts without (-) or with (+) 3 mM 6-DMAP. 'No' is insoluble fraction without sperm nuclei added. LSS, *Xenopus* low speed egg extract. **f**, XCDT1 is released from chromatin by salt wash. Western blotting of proteins eluted from pre-replicative chromatin isolated with the indicated concentration of KCl. **g**, Diagram of formation of pre-replication complexes in *X. laevis*.

later times was not observed (Fig. 2a, 5–15 min). Instead, the level of chromatin-bound XCDC1 started to decline at the time of activation of DNA synthesis (Fig. 3b, 30 min) and dropped to background levels thereafter (Fig. 3b, 50–80 min). Inhibition of DNA synthesis with aphidicolin, a potent inhibitor of DNA polymerase- α , during the same time course, stabilized chromatin-bound XCDC1 ('+aphi' in Fig. 3c). XCDC1 was only transiently stabilized on chromatin in the presence of aphidicolin, compared with XMCM4, which remained bound as described¹⁶. Collectively, these data indicate that XCDC1 is destabilized from chromatin upon initiation of DNA synthesis and that its function may be required before the DNA polymerases execution step.

We then determined whether the *Xenopus* cdt1 protein may be involved in some steps of the formation of pre-replication complexes. It has been shown that this process occurs by the sequential assembly of ORC, Cdc6 and MCM proteins^{4,17,18}. Cdc6, which is required to load MCM proteins onto chromatin, binds only if ORC is present but its binding is independent of MCMs. Chromatin assembled in XCDC1-depleted extracts is fully loaded with XCDC6 protein (Fig. 4a); however, this chromatin contains very low levels of XMCM2–7 proteins, similar to those observed on MCM-depleted chromatin (Fig. 4a). Analysis of chromatin formed in XORC1-depleted extracts shows that XCDC1 is not bound (Fig. 4b), whereas chromatin formed in XCDC1-depleted extracts contains ORC1 (Fig. 4b). Thus the association of XCDC1 with chromatin requires the presence of the XORC complex. These results also indicate that despite ORC and CDC6 being bound, accumulation of XMCM2–7 proteins on chromatin does not occur in the absence of XCDC1. XCDC1 is not physically associated with XMCM proteins in *Xenopus* egg extracts (Fig. 2a and data not shown) suggesting that XCDC1 may bind chromatin independently of MCM proteins. If so, XCDC1 should be able to assemble on chromatin in extracts lacking XMCM proteins and this pre-loaded chromatin should be competent to replicate when transferred to XCDC1-depleted extracts (Fig. 4c). Chromatin-transfer experiments (Fig. 4d) show that chromatin that is pre-formed in extracts lacking MCM proteins ('MCM-' in Fig. 4d) replicates when it is transferred to XCDC1-depleted extracts that contain MCM proteins. This result indicates that XCDC1, as XCDC6 (ref. 4), can assemble on chromatin in the absence of MCM proteins. These experiments also imply that XCDC1 and XMCM2–7 proteins do not need to form a complex just before binding to chromatin. As expected, the reverse was not true (Fig. 4d); that is, chromatin formed in extracts lacking XCDC1 ('XCDC1-' in Fig. 4d, see Fig. 4b) was not competent to replicate in MCM-depleted extracts, confirming that XCDC1 is essential for the loading of MCM proteins onto chromatin.

In the presence of 6-DMAP, an inhibitor of the licensing reaction¹⁹, the XMCM3 protein does not associate with chromatin^{11,19,4} (Fig. 4e). However, the XCDC1 protein is efficiently bound, at apparently higher levels than those found in control extracts (Fig. 4e), similar to XORC1 and XCDC6 (refs 4, 17). Chromatin pre-formed in egg extracts treated with 6-DMAP is competent to replicate in XCDC1-depleted extracts (Fig. 4d), suggesting that XCDC1 bound to 6-DMAP-treated chromatin may be functional. After formation of pre-replication complexes, XCDC1 can be removed from chromatin by salt wash, without displacing XMCM2 (Fig. 4f), suggesting that XCDC1 is not required for the retention of XMCM2 on chromatin. XORC1 and XCDC6 are also removed from licensed chromatin by a similar treatment without displacing MCM3 (ref. 20). Collectively, these data show that XCDC1 is required for the assembly of MCM proteins onto chromatin but its loading is independent of XMCMs.

We conclude that *Xenopus* CDT1 is a replication factor that is essential in loading MCM proteins onto pre-replicative chromatin, a step of pre-replication complexes assembly. This step is likely to be accomplished in combination with XCDC6 (Fig. 4g), itself required for the chromatin association of XMCMs (refs 4, 20). Both XCDC1

and XCDC6 require XORC for chromatin binding; however, XCDC6 binds in the absence of XCDC1, suggesting that both proteins assemble on chromatin independently of each other. XCDC1 is functionally bound to unlabeled, 6-DMAP-treated chromatin, indicating that it may be distinct from the activity (RLF-B) inhibited by 6-DMAP (ref. 19). XCDC1 is destabilized from chromatin once pre-replication complexes have formed; thus, it may possibly be involved in the block to re-initiation in S phase. XCDC1 appears to be conserved in evolution, except in budding yeast. Whether or not this reflects a divergence in the organization of the genome for origin recognition remains an open issue. It has been suggested that Cdc6 shares similarity with a group of prokaryotic and eukaryotic nucleotide-dependent loading factors²¹. Both *Xenopus* and fission yeast cdt1 are unrelated to this protein family, indicating that cdt1 is a new component of the licensing system. □

Methods

Plasmids and proteins expression

XCDC1 was isolated from a subtraction cDNA library made from messenger RNAs enriched in the early *Xenopus* embryo⁸, sub-cloned into the Bluescript vector (Stratagene) digested with *EcoRI* to generate pBSmat5 and sequenced on both DNA strands by the dye termination method. A 6 × His-tagged XCDC1 fusion protein (amino acids 141–621) was made by subcloning into the 6 × His-tag vector pRSET-B (Invitrogen) to generate p6HΔXCDC1. The recombinant protein was expressed in *Escherichia coli* and purified under denaturing conditions on a Ni²⁺-NTA column following the instructions of the supplier (Qiagen). After purification, the protein was renatured *in vitro* as described²², dialysed against Xb (100 mM KCl, 0.1 mM CaCl₂, 50 mM sucrose, 10 mM HEPES-KOH pH 7.7, 2 mM MgCl₂, 5 μg ml⁻¹ leupeptin, aprotinin, pepstatin), concentrated in microcon-30 (Amicon) and stored at -20 °C in 10% glycerol. *In vitro* translation was done using a reticulocyte system (TNT, Promega).

Xenopus egg extracts

Low speed interphasic extracts (LSS)²³, CSF *Xenopus* egg extracts²⁴, high speed *Xenopus* egg extracts¹² and 6-DMAP-treated extracts¹⁴ were made as described. Glycerol (to 3%) was added to the extracts, which were frozen in liquid nitrogen and stored at -80 °C. Upon thawing, extracts were supplemented with an ATP regeneration system (10 mM creatine phosphate, 10 μg ml⁻¹ creatine kinase, 1 mM ATP, 1 mM MgCl₂).

Immunological techniques

XCDC1 polyclonal antibodies were raised in rabbits against recombinant 6HisΔXCDC1 protein further purified by SDS-PAGE. MCM proteins were detected with an antibody that recognizes a common epitope in the *Xenopus* MCM proteins (MCM signature)²⁵. The XCDC6 protein was detected with an antibody raised against fission yeast CDC6 homologue that strongly crossreacts with purified XCDC6 (obtained from S. Cotterill, MCRI, UK). XORC1 antibody has been described¹⁷. Depletion of *Xenopus* egg extracts was carried out as described^{17,25}. Depleted extracts were supplemented with an ATP regeneration system, sperm nuclei (4 ng μl⁻¹) or ssM13 DNA (6 ng μl⁻¹); [α-³²P]dATP and DNA replication was monitored for the time required at 23 °C. For rescue with proteins eluted from XCDC1 antibody, beads were extensively washed with Xb and eluted with 1 volume of Xb/1M NaCl. Eluates were dialysed against Xb and concentrated in microcon-30.

Chromatin purification

Demembrated sperm nuclei, prepared as described²³, were assembled in activated *Xenopus* eggs supernatants (LSS) as described^{17,25}. For salt extraction of chromatin, KCl was included in the extraction buffer at the indicated concentrations. 6-DMAP chromatin was made as described¹⁷.

Received 19 October 1999; accepted 3 February 2000.

- Cocker, J. H., Piatti, S., Santocanale, C., Nasmyth, K. & Diffley, J. F. X. An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature* **379**, 180–182 (1996).
- Tanaka, T., Knapp, D. & Nasmyth, K. Loading of an MCM protein onto DNA replication origins is regulated by Cdc6p and CDKs. *Cell* **90**, 649–660 (1997).
- Aparicio, O. M., Weinstein, D. M. & Bell, S. P. Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell* **91**, 59–69 (1997).
- Coleman, T. R., Carpenter, P. B. & Dunphy, G. The *Xenopus* Cdc6 protein is essential for the initiation of a single round of DNA replication in cell-free extracts. *Cell* **87**, 53–63 (1996).
- Ishimi, Y. A DNA helicase activity is associated with an MCM4, -6 and -7 protein complex. *J. Biol. Chem.* **272**, 24508–24513 (1997).
- Liang, C. & Stillman, B. Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants. *Genes Dev.* **11**, 3375–3386 (1997).
- Hofmann, J. F. X. & Beach, D. cdt1 is an essential target of the Cdc10/Scf1 transcription factor: requirement for DNA replication and inhibition of mitosis. *EMBO J.* **13**, 425–434 (1994).
- Moreau, J., Ioualen, N. & Méchali, M. Isolation of cDNAs from maternal mRNAs specifically present during early development. *Mol. Rep. Dev.* **41**, 1–7 (1995).

9. Fernandez Sarabia, M. J., McNerny, C., Harris, P., Gordon, C. & Fantes, P. The cell cycle genes *cdc22+* and *suc22+* of the fission yeast *Schizosaccharomyces pombe* encode the large and small subunits of ribonucleotide reductase. *Mol. Gen. Genet.* **238**, 241–251 (1993).
10. Kelly, T. J. *et al.* The fission yeast *cdc18+* gene product couples S phase to START and mitosis. *Cell* **74**, 371–382 (1993).
11. Carpenter, P. B., Mueller, P. R. & Dunphy, W. Role for a *Xenopus* Orc2-related protein in controlling DNA replication. *Nature* **379**, 357–360 (1995).
12. Mechali, M. & Harland, R. M. DNA synthesis in a cell-free system from *Xenopus* eggs: priming and elongation on single-stranded DNA *in vitro*. *Cell* **30**, 93–101 (1982).
13. Almouzni, G. & Mechali, M. Assembly of spaced chromatin promoted by DNA synthesis in extracts from *Xenopus* eggs. *EMBO J.* **7**, 665–672 (1988).
14. Blow, J. J. Preventing re-replication of DNA in a single cell cycle: evidence for a replication licensing factor. *J. Cell Biol.* **122**, 993–1002 (1993).
15. Coue, M., Amariglio, F., Maiorano, D., Bocquet, S. & Mechali, M. Evidence for different MCM subcomplexes with differential binding to chromatin in *Xenopus*. *Exp. Cell Res.* **245**, 282–289 (1998).
16. Coué, M., Kearsey, S. E. & Mechali, M. Chromatin binding, nuclear localization and phosphorylation of *Xenopus* *cdc21* are cell-cycle dependent and associated with the control of initiation of DNA replication. *EMBO J.* **15**, 1085–1097 (1996).
17. Rowles, A. *et al.* Interaction between the origin recognition complex and the replication licensing system in *Xenopus*. *Cell* **87**, 287–296 (1996).
18. Romanowski, P., Madine, M. A., Rowles, A., Blow, J. J. & Laskey, R. A. The *Xenopus* origin recognition complex is essential for DNA replication and MCM binding to chromatin. *Curr. Biol.* **6**, 1416–1425 (1996).
19. Chong, J. P., Mahbubani, H. M., Khoo, C. -Y. & Blow, J. J. Purification of an MCM-containing complex as a component of the DNA replication licensing system. *Nature* **375**, 418–421 (1995).
20. Rowles, A., Tada, S. & Blow, J. J. Changes in association of the *Xenopus* origin recognition complex with chromatin on licensing of replication origins. *J. Cell Sci.* **112**, 2011–8 (1999).
21. Perkins, G. & Diffley, J. F. Nucleotide-dependent prereplicative complex assembly by Cdc6p, a homolog of eukaryotic and prokaryotic clamp-loaders. *Mol. Cell* **2**, 23–32 (1998).
22. Vuillard, L., Rabilloud, T. & Goldberg, M. E. Interactions of non-detergent sulfolipids with early folding intermediates facilitate *in vitro* protein renaturation. *Eur. J. Biochem.* **256**, 128–135 (1998).
23. Menut, S., Lemaître, J. M., Hair, A. & Mechali, M. in *Advances in Molecular Biology: A Comparative Methods Approach to the Study of Oocytes and Embryos* (ed. Richter, J. D.) 198–226 (Oxford Univ. Press, 1999).
24. Murray, A. W. Cell cycle extracts. *Methods Cell Biol.* **36**, 581–605 (1991).
25. Maiorano, D., Lemaître, J.-M. & Mechali, M. Step-wise, regulated chromatin assembly of MCM2-7 proteins. *J. Biol. Chem.* (in the press).

Acknowledgements

We thank S. Bocquet for technical assistance in the raising of antibodies; S. Cotterill for the cdc6 antibody; J. Blow and A. Rowles for the ORC1 antibody; and C. Jaulin, D. Fisher and J.-M. Lemaître for helpful technical tips. This work has been supported by grants from ARC and Ligue Nationale contre le Cancer. D. M. was supported by an EC fellowship (Biomedicine and Health) and a grant from Fondation pour la Recherche Médicale.

Correspondence and requests for materials should be addressed to M. M. (e-mail: mechali@igh.cnrs.fr).

The Cdt1 protein is required to license DNA for replication in fission yeast

Hideo Nishitani^{*†‡}, Zoi Lygerou^{*‡§}, Takeharu Nishimoto[†] & Paul Nurse^{*}

^{*} Imperial Cancer Research Fund, 44, Lincoln's Inn Fields, London WC2A 3PX, UK

[†] Graduate School of Medical Science, Kyushu University, Fukuoka 812-8582, Japan

[‡] These authors contributed equally to this work.

To maintain genome stability in eukaryotic cells, DNA is licensed for replication only after the cell has completed mitosis, ensuring that DNA synthesis (S phase) occurs once every cell cycle¹. This licensing control is thought to require the protein Cdc6 (Cdc18 in fission yeast) as a mediator for association of minichromosome maintenance (MCM) proteins with chromatin^{2–10}. The control is overridden in fission yeast by overexpressing Cdc18 (ref. 11) which leads to continued DNA synthesis in the absence of mitosis¹². Other factors acting in this control have been

postulated¹³ and we have used a re-replication assay to identify Cdt1 (ref. 14) as one such factor. Cdt1 cooperates with Cdc18 to promote DNA replication, interacts with Cdc18, is located in the nucleus, and its concentration peaks as cells finish mitosis and proceed to S phase. Both Cdc18 and Cdt1 are required to load the MCM protein Cdc21 onto chromatin at the end of mitosis and this is necessary to initiate DNA replication. Genes related to Cdt1 have been found in Metazoa and plants (A. Whitaker, I. Roysman and T. Orr-Weaver, personal communication), suggesting that the cooperation of Cdc6/Cdc18 with Cdt1 to load MCM proteins onto chromatin may be a generally conserved feature of DNA licensing in eukaryotes.

Cdc6/18 has a central role in regulating S-phase onset in a number of eukaryotes¹⁵. We employed a re-replication assay in fission yeast to identify new factors that cooperate with Cdc18 to induce DNA synthesis. Increased expression of Cdc18 induces fission yeast cells to undergo continued DNA synthesis in the absence of mitosis, leading to re-replication¹². When Cdc18 was expressed in cells at increasing levels (low, L; medium, M; and high, H), a corresponding increase in the extent of re-replication was observed (Fig. 1a, b). There was hardly any re-replication in strain L, whereas an average DNA content of 6C and 16C was observed in strains M and H, respectively. We reasoned that co-expression of a factor that cooperates with Cdc18 might induce the low-expressing strain L to re-replicate more effectively. The fission yeast Cdt1 protein is a good candidate for a Cdc18 partner as it is under the same transcriptional regulation as Cdc18 (ref. 14) and, like Cdc18 (ref. 11), is required both for S-phase onset and for the DNA replication checkpoint¹⁴. Overexpression of Cdt1 did not promote re-replication on its own (data not shown). However, co-expression of Cdt1 in strain L generated cells with an average DNA content of 16C, similar to the level of re-replication observed in strain H and significantly higher than the level seen in the medium-expressing strain M (Fig. 1c, d). We saw no enhancement of DNA synthesis when Cdc30 (SpOrc1; refs 16, 17), Hsk1 (ref. 18; a CDC7 homologue) or Cdc22 (ref. 19; the large subunit of ribonucleotide reductase) were co-expressed with Cdc18 (data not shown). We conclude that Cdt1 expression enhances the effect of Cdc18 to promote continuing DNA synthesis, indicating that Cdc18 and Cdt1 cooperate to induce DNA replication.

To determine whether Cdt1 and Cdc18 can physically interact, Cdt1 was tagged with glutathione S-transferase (GST–Cdt1) and Cdc18 with the myc epitope (myc–Cdc18), and both fusion proteins were co-expressed in fission yeast. As shown in Fig. 2a, precipitation of GST–Cdt1 (and not the control GST protein) specifically co-precipitated myc–Cdc18 (compare lanes 3 and 4, lower panel). The reciprocal experiment is shown in Fig. 2b; immunoprecipitation of myc–Cdc18 specifically co-precipitates GST–Cdt1. The association between Cdt1 and Cdc18 was confirmed using cells expressing untagged Cdt1 and Cdc18 (Fig. 2c); immunoprecipitation with anti-Cdt1 antibodies specifically brings down Cdc18 (lane 4). To map the domain of Cdc18 required for Cdt1 interaction, the amino-terminal and carboxy-terminal regions of Cdc18 were tagged with the haemagglutinin (HA) epitope and independently expressed together with either GST–Cdt1 or GST alone as a control. As shown in Fig. 2d, the C terminus but not the N terminus of Cdc18 is specifically co-precipitated with GST–Cdt1 (lane 6). We therefore conclude that Cdc18 can form *in vivo* complexes with Cdt1, primarily through its C terminus. We have previously shown that it is the C terminus of Cdc18 which induces re-replication and activates the DNA replication checkpoint, whereas the N terminus associates with the Cdc2 protein kinase²⁰. The re-replication induced by the C terminus of Cdc18 is also enhanced by co-expression of Cdt1, to a level similar to that seen with the full-length protein (Fig. 2e).

Cdc18 is located in the nucleus, and its protein levels vary periodically through the cell cycle^{12,21}. To determine whether Cdt1

[§] Present address: Laboratory of General Biology, School of Medicine, University of Patras, 26110, Rio, Patras, Greece.

behaves similarly, we investigated the levels and subcellular localization of Cdt1 through the cell cycle. We found that Cdt1 protein levels fluctuated as cells progressed through a synchronous cell cycle, peaking at the end of mitosis, in parallel with Cdc18 protein levels (Fig. 2f). Both Cdc18 and Cdt1 localized to the nuclei of binucleate cells that have just completed mitosis and are in G1 or early S phase (Fig. 2g). Either no, or very weak, nuclear staining was observed in septated cells which were further on in S phase, or in G2 cells. These observations were confirmed using synchronous cultures (data not shown). Cdt1 and Cdc18 are therefore present in the cell nucleus after mitosis when DNA licensing takes place.

DNA licensing is believed to require chromatin association of the MCM proteins²⁻⁵, and in budding yeast and *Xenopus* this association depends on the Cdc6 protein⁶⁻¹⁰, the homologue of Cdc18 in these organisms. To confirm MCM protein association with chromatin in fission yeast, and establish the requirement for Cdc18 and

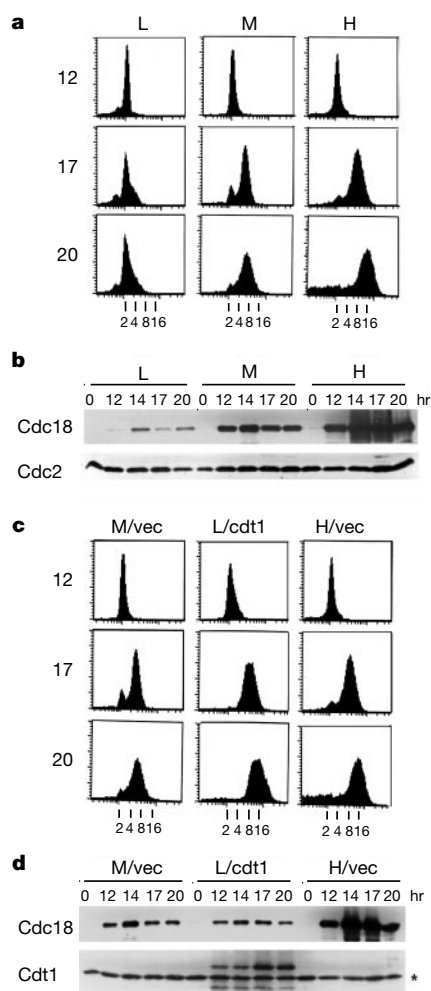


Figure 1 Cdt1 enhances the ability of Cdc18 to induce continuing DNA synthesis. **a**, **b**, Three strains were constructed which express Cdc18, under the inducible nmt promoter, at different levels (L, low; M, medium; H, high). The promoter was derepressed and cells collected at the indicated time points (in h). The DNA content was analysed by fluorescence-activated cell sorting (FACS) analysis (**a**) and the level of Cdc18 produced was monitored by western blotting (**b**). Cdc2 protein levels are shown as loading control. **c**, **d**, Strain L was transformed with plasmid pREP4-Cdt1, in which Cdt1 is under the nmt promoter, and expression of both Cdc18 and Cdt1 was induced. As controls, strains M and H transformed with vector alone were used. DNA content was measured by FACS (**c**) and levels of both Cdc18 and Cdt1 were assessed by western blotting (**d**). Cdt1 overexpression led to a small increase in Cdc18 protein levels, perhaps due to stabilization of the Cdc18 protein in rereplicating cells. The band marked with an asterisk in **d** is a non-specific band that serves as loading control.

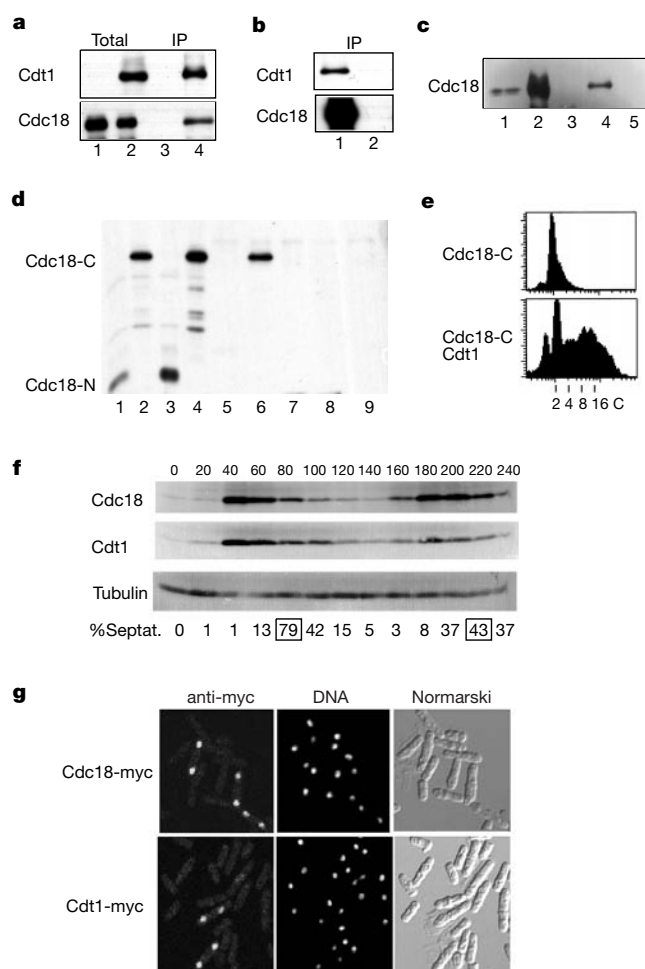


Figure 2 Cdt1 can form a complex with Cdc18, and accumulates in the cell nucleus in G1. **a**, Cell extracts were prepared from cells co-expressing myc-Cdc18 (in pRMH41) and GST (lanes 1 and 3) or GST-Cdt1 (lanes 2 and 4) (both in pREP4) and the GST proteins precipitated with glutathione beads. Cell extracts (lanes 1 and 2) and precipitates (lanes 3 and 4) were subjected to western blotting with anti-Cdt1 (upper panel) or anti-Cdc18 antibodies (lower panel). 5 μ g cell extract and the precipitate from 25 μ g extract, and 1 μ g cell extract and the precipitate from 125 μ g extract, were loaded for the Cdt1 and Cdc18 blots, respectively. **b**, Myc-Cdc18 in the cell extract used in **a** lane 2, was precipitated with anti-myc (lane 1) or anti-HA antibodies (lane 2), and the precipitates blotted with anti-Cdt1 (upper panel) or anti-Cdc18 antibodies (lower panel). Precipitates from 125 μ g and 25 μ g of extract were loaded for the Cdt1 blot and the Cdc18 blot (lower panel), respectively. **c**, Untagged Cdc18 and Cdt1 were co-expressed using pREP41-Cdc18 and pREP4-Cdt1. Immunoprecipitation was performed using anti-Cdc18 (lane 2) or anti-Cdt1 antibodies (lane 4) or pre-immune serum for Cdc18 (lane 3) or Cdt1 (lane 5). 25 μ g of total cell extract (lane 1) and precipitates from 125 μ g extract (lanes 2–5) were western blotted using anti-Cdc18 antibodies. **d**, N terminus of Cdc18 (residues 1–141 in pRHA41) tagged with HA was co-expressed with GST-Cdt1 (lanes 1 and 5) or GST alone (lanes 3 and 7), both in pREP4X. Similarly, the C terminus of Cdc18 (residues 150 to end in pRHA41) tagged with HA was co-expressed with GST-Cdt1 (lanes 2, 6 and 9) or GST alone (lanes 4 and 8). 2.5 μ g cell extract (lanes 1–4) and precipitates from 250 μ g extract using glutathione beads (lanes 5–8) or BSA beads as control (lane 9) were western blotted with anti-HA antibodies. **e**, Cdc18-C (in pREP81) was expressed alone (upper panel) or with Cdt1 (in pREP42) (lower panel) and the DNA content measured after 20 h expression. **f**, *cdc25-22* cells were synchronized at the G2–M transition by incubating for 4 h at the non-permissive temperature and released into a synchronous cell cycle. Samples were taken every 20 min and the levels of Cdc18 (top), Cdt1 (middle) and α -tubulin (bottom, loading control) were assessed by western blotting. Per cent septation is indicated at the bottom. Initiation of DNA replication immediately precedes the peak of septation. **g**, Cells expressing Cdc18-myc (upper panel) or Cdt1-myc (lower panel), both under their endogenous promoters, were fixed with methanol and stained with anti-myc antibody (left), propidium iodide (middle) or visualized by Nomarski optics (right). Only binucleate cells show strong nuclear staining for either protein.

Cdt1 for this association, we raised antibodies against Cdc21 (ref. 22; an MCM4 homologue) and developed a fractionation protocol (Fig. 3a) which enriches for chromatin-associated components^{7,8,23}. In the total cell extracts and detergent-soluble protein fractions, Cdc21 protein levels remained constant through the cell cycle. However, the appearance of Cdc21 in the chromatin-enriched preparation was periodic through the cell cycle. When synchronized cells completed mitosis, Cdc21 became associated with chromatin and this association persisted through G1 into S phase (Fig. 3b, left bottom panel, Fig. 3c). It decreased to a low level during G2 and reappeared after the second mitosis. SpOrc1 protein levels in total,

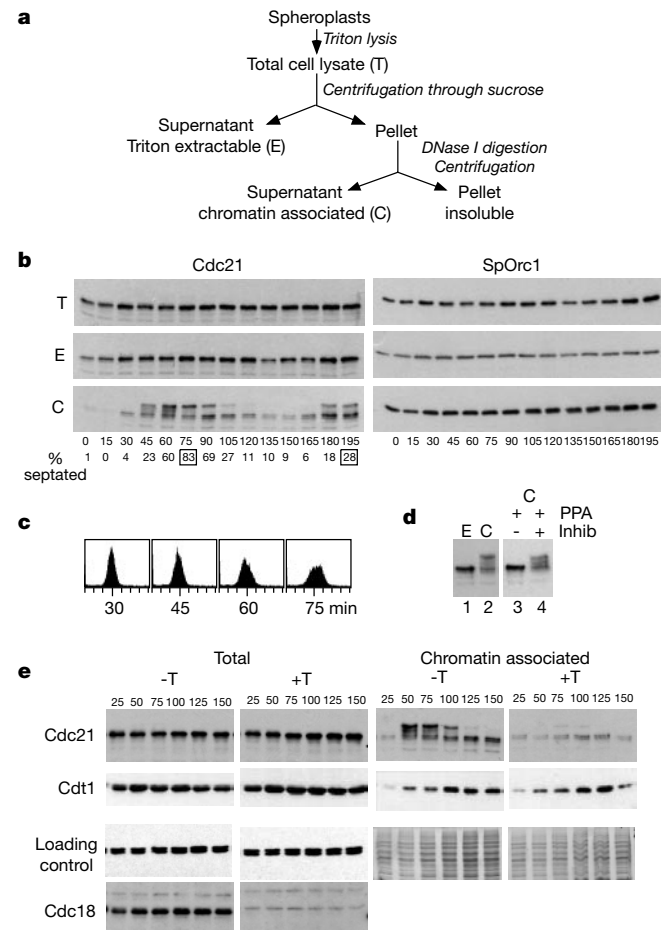


Figure 3 Cdc21 associates with chromatin as cells prepare for S phase in a Cdc18-dependent manner. **a**, Flowchart of chromatin association assay. **b**, **c**, *cdc25-22*HA-Orc1 cells were blocked at the G2–M transition and released into a synchronous cell cycle. Samples collected every 15 min were fractionated as shown in **a**. The total cell lysate (T), Triton extractable (E) and chromatin-associated fractions (C) were analysed by western blotting using anti-Cdc21 (left) or anti-HA antibodies (right). A higher exposure is shown for Cdc21 (C). We estimate that around 20% of the total Cdc21 is found in the chromatin-enriched fraction in G1. Per cent septated cells are indicated at the bottom. FACS analysis (**c**) shows when initiation of DNA replication occurs. **d**, The Triton extractable (lane 1) and chromatin-associated fraction (lanes 2–4) from the 60 min time point of **b** were western blotted for Cdc21 without prior treatment (lane 2) or after incubation with calf intestinal alkaline phosphatase in the absence (lane 3) or presence (lane 4) of phosphatase inhibitors. **e**, *cdc25-22* cells expressing Cdc18 under the control of the thiamine-repressible nmt promoter as their only functional copy were blocked at G2–M for 4 h at 36 °C in the absence (–T) or presence (+T) of thiamine and then released to 25 °C (always – or +T). Samples were taken every 25 min and fractionated as in **a**. Total cell lysate and chromatin-associated fractions were western blotted using anti-Cdc21, anti-Cdt1, and anti-Cdc18 antibodies as shown. A higher exposure is shown for the chromatin-associated fraction of Cdc21. As loading controls, an anti- α -tubulin western blot is shown for the total cell lysate, and a Coomassie-stained gel of the chromatin-associated fractions.

soluble and chromatin fractions were constant and served as loading controls. A significant fraction of the chromatin-associated Cdc21 migrated slower than the detergent-soluble protein (Fig. 3d, lanes 1 and 2). This was due to phosphorylation of the chromatin-associated Cdc21 (Fig. 3d, lanes 3 and 4). We used a strain in which Cdc18 was under the control of the thiamine-repressible low strength nmt promoter to test the effect of Cdc18 depletion on Cdc21 chromatin association. Cells depleted of Cdc18 at the G2–M transition completed mitosis normally but failed to initiate DNA replication (data not shown). In the absence of Cdc18, there was no increase in Cdc21 chromatin association (Fig. 3e) or phosphorylation (data not shown) as cells completed mitosis and entered G1. This establishes that Cdc18 is required for the chromatin association and phosphorylation of Cdc21 at the G1–S transition. Cdt1 was also found to be present in the chromatin-enriched fraction during the G1 and S phases of the cell cycle, but in contrast to Cdc21 there was no requirement for Cdc18 to bring about this association (Fig. 3e).

We next investigated the effects of Cdt1 depletion on DNA replication and MCM association with chromatin. To find out whether Cdt1 is primarily required for early or late S phase, cells in which Cdt1 is under the control of the nmt promoter were arrested with hydroxyurea at the beginning of S phase. Thiamine

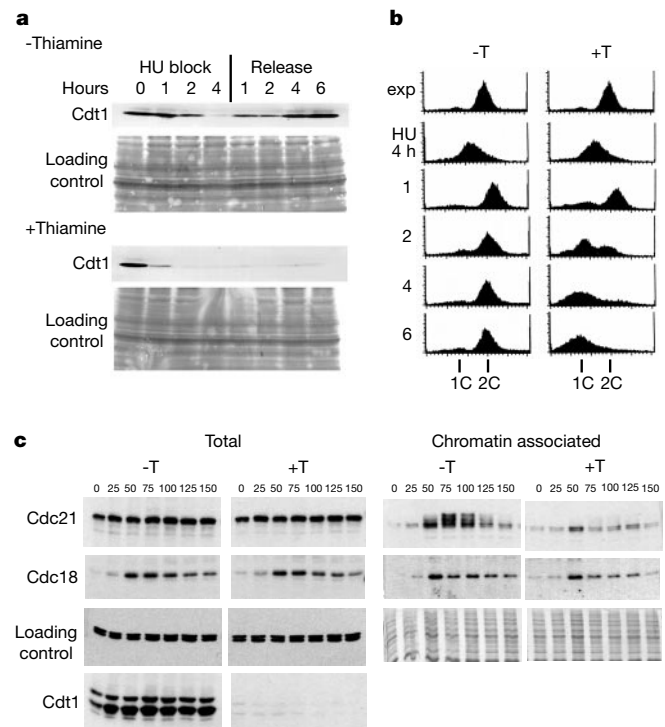


Figure 4 Cdt1 depletion blocks the initiation of DNA replication by inhibiting the association of Cdc21 with chromatin. **a**, **b**, Cells expressing Cdt1 under the control of the thiamine repressible nmt promoter as their only functional copy (nmt-cdt1) were arrested at early S phase with hydroxyurea (HU), in the absence or presence of thiamine (–T, +T). After 4 h, HU was washed off and cells were released, always –T or +T. A western blot with anti-Cdt1 antibodies is shown in **a**. Coomassie staining serves as loading control. Cdt1 levels decrease after 4 h in HU even in the absence of thiamine, indicating that Cdt1 is unstable in HU blocked cells. DNA content at the indicated times (h) is shown in **b**. **c**, *cdc25-22* nmt-cdt1 cells were arrested with HU for 6 h at 25 °C without or with thiamine (–T or +T). After removal of the HU, cells were shifted to 36 °C for 3 h to arrest them at the G2–M transition and then released into a synchronous cell cycle, always in the presence or absence of thiamine. Samples were collected every 25 min and processed as in Fig. 3a. The total cell lysate and the chromatin-associated fraction were western blotted and visualized with anti-Cdc21, anti-Cdc18 and anti-Cdt1 antibodies as indicated while an α -tubulin western blot and Coomassie staining serve as loading controls for the total lysate and chromatin-extractable fractions, respectively.

was added to half the culture at the same time as the hydroxyurea to repress Cdt1 expression, and a western blot showed that Cdt1 levels dropped by over 90% within two hours (Fig. 4a). After washing off the hydroxyurea, both the cells expressing Cdt1 (Fig. 4b, left panels) and the Cdt1 depleted cells (Fig. 4b, right panels) completed DNA replication within one hour. Thus, Cdt1 depletion does not prevent the later stages of DNA replication. However, cells lacking Cdt1 were unable to undergo a second round of DNA replication and accumulated in G1. We conclude that Cdt1 is required for the initial stages of DNA replication. Cells lacking Cdt1 subsequently proceeded inappropriately into mitosis, generating a 'cut' phenotype (data not shown), verifying that Cdt1 is required for the checkpoint control inhibits mitosis when S phase has not taken place¹⁴. To determine whether Cdt1 depleted cells were defective in the association of Cdc21 with chromatin, cells were released from a hydroxyurea block, synchronized at the G2/M transition, and samples taken as cells completed mitosis and progressed to S phase. Like Cdc18 depleted cells, Cdt1 depleted cells were defective in both the chromatin association and phosphorylation of Cdc21 (Fig. 4c). In contrast, Cdc18 could still associate with chromatin in the absence of Cdt1, although possibly at a reduced level. We conclude that Cdt1 is required to promote the association of Cdc21 with chromatin in preparation for S phase.

We have identified Cdt1 as being required to license DNA for replication after exit from mitosis. It can associate with Cdc18 in cells, it accumulates within the nucleus, associates with chromatin and is expressed periodically during the cell cycle, peaking from mitotic exit to the onset of S phase. Elevated Cdt1 expression promotes the ability of Cdc18 to bring about continued DNA synthesis, and the absence of either Cdt1 or Cdc18 blocks the association of Cdc21 (MCM4) with chromatin and prevents the initiation of DNA replication. We propose that Cdt1 acts cooperatively with Cdc18 during the G1 phase of the cell cycle, bringing about the construction of pre-initiation complexes on replication origins by loading MCM proteins onto chromatin associated with the origin recognition complex (ORC). Genes related to Cdt1 have been found in *Drosophila*, human, mouse, zebrafish, *Caenorhabditis elegans* and *Arabidopsis*, although not yet in budding yeast, and mutant analysis of the Cdt1 related gene in *Drosophila* indicates that it is required for DNA replication (A. Whitaker, I. Roysman and T. Orr-Weaver, personal communication). This suggests that the licensing function of Cdt1 for S-phase onset may be conserved in multicellular eukaryotes. Its apparent absence from budding yeast may be due to a greater divergence of the primary sequence of Cdt1 in this organism, or it may reflect differences in origin organization or chromatin structure. □

Methods

Strains and cell culture conditions

All strains were derived from the wild-type 972 h⁻ or 975 h⁺ and standard media and growth conditions were used^{24,25}. To construct the nmt cdc18 integrants for re-replication experiments, plasmid pJK-nmt-cdc18 (ref. 12) was used to transform a *leu1-32 ura4D-18* strain, and three integrants expressing Cdc18 to different levels were isolated. The Cdt1 deletion strain was constructed as described¹⁴. The strains used for switch-off experiments were: *cdc25-22 cdc18::ura4⁺ ura4D-18 leu1-32* [pREP81-cdc18] and *cdc25-22 cdt1::ura4⁺ ura4D-18 leu1-32* [pREP81-cdt1]. Myc-tagged Cdc18 or Cdt1 were constructed by inserting the myc epitope at the C terminus of the corresponding endogenous gene.

Plasmids

pREP41-myc-Cdc18 was constructed by inserting the *SalI*-*BamHI* *cdc18⁺* cDNA to plasmid pRMH41. Plasmids expressing either the C terminus (150 to end) or N terminus (1-141) of Cdc18 tagged with HA were constructed by inserting the polymerase chain reaction (PCR) products into pRHA41. The *cdt1⁺* gene was amplified by PCR using genomic DNA as a template, cloned and sequenced. We found one base difference to the published sequence (from T to C at position 53). This difference was confirmed by sequencing the *cdt1⁺* gene isolated from a genomic library.

Antibodies and immunodetection

The anti-Cdc18 antibody has been described¹². For cdt1 antibody production, His-tagged Cdt1 (residue 201-end in vector pQE) purified from *Escherichia coli* was used for rabbit

immunizations. Anti-Cdc21 antibodies were produced against the C terminus of Cdc21 as described²².

Total cell extract used for western blotting was prepared by glass beads²⁴. For pull-down experiments, cells were spheroplasted, lysed in HB buffer¹⁷ and spun down at 100,000g for 30 min. The supernatant was mixed with antibodies or glutathione beads. Immunofluorescence analysis was done with methanol fixed cells processed as described²⁴.

Chromatin association assay

Spheroplast preparation and lysis was performed as described²³. The total cell extract was overlaid on a 30% sucrose cushion, and spun at 16,000g for 15 min. The material remaining soluble was removed (Triton extractable fraction) while the chromatin-associated proteins were released from the washed and repelleted chromatin by incubation with 277 U DNase I (Sigma) for 10 min at 25 °C in a buffer containing 20 mM Hepes pH 7.9, 1.5 mM magnesium acetate, 50 mM potassium acetate, 10% glycerol, 0.5 mM DTT, 150 mM NaCl, protease and phosphatase inhibitors, and were then analysed by SDS-polyacrylamide gel electrophoresis.

Received 8 November 1999; accepted 3 February 2000.

- Blow, J. J. & Laskey, R. A. A role for the nuclear envelope in controlling DNA replication within the cell cycle. *Nature* **332**, 546-548 (1988).
- Kubota, Y., Mimura, S., Nishimoto, S., Takisawa, H. & Nojima, H. Identification of the yeast MCM3-related protein as a component of *Xenopus* DNA replication licensing factor. *Cell* **81**, 601-609 (1995).
- Chong, J., Mahbubani, H., Khoo, C. & Blow, J. Purification of an MCM-containing complex as a component of the DNA replication licensing system. *Nature* **375**, 418-421 (1995).
- Madine, M., Khoo, C., Mills, A. D. & Laskey, R. A. MCM3 complex required for cell cycle regulation of DNA replication in vertebrate cells. *Nature* **375**, 421-424 (1995).
- Rowles, A. *et al.* Interaction between the origin recognition complex and the replication licensing system in *Xenopus*. *Cell* **87**, 287-296 (1996).
- Coleman, T. R., Carpenter, P. B. & Dunphy, W. G. The *Xenopus* Cdc6 protein is essential for the initiation of a single round of DNA replication in cell-free extracts. *Cell* **87**, 53-63 (1996).
- Donovan, S., Harwood, J., Drury, L. S. & Diffley, J. F. Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proc. Natl Acad. Sci. USA* **94**, 5611-5616 (1997).
- Liang, C. & Stillman, B. Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in *cdc6* mutants. *Genes Dev.* **11**, 3375-3386 (1997).
- Aparicio, O. M., Weinstein, D. M. & Bell, S. P. Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell* **91**, 59-69 (1997).
- Tanaka, T., Knapp, D. & Nasmyth, K. Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. *Cell* **90**, 649-660 (1997).
- Kelly, T. J. *et al.* The fission yeast *cdc18* gene product couples S-phase to start and mitosis. *Cell* **74**, 371-382 (1993).
- Nishitani, H. & Nurse, P. p65^{cdc18} plays a major role controlling the initiation of DNA replication in fission yeast. *Cell* **83**, 397-405 (1995).
- Tada, S., Chong, J. P. J., Mahbubani, H. M. & Blow, J. J. The RLF-B component of the replication licensing system is distinct from Cdc6 and functions after Cdc6 binds to chromatin. *Curr. Biol.* **9**, 211-214 (1999).
- Hofmann, J. F. X. & Beach, D. *cdt1* is an essential target of the Cdc10/Sct1 transcription factor: requirement for DNA replication and inhibition of mitosis. *EMBO J.* **13**, 425-434 (1994).
- Dutta, A. & Bell, S. P. Initiation of DNA replication in eukaryotic cells. *Annu. Rev. Cell. Dev. Biol.* **13**, 293-332 (1997).
- Muzi-Falconi, M. & Kelly, T. J. Orp1, a member of the Cdc18/Cdc6 family of S-phase regulators, is homologous to a component of the origin recognition complex. *Proc. Natl Acad. Sci. USA* **92**, 12475-12479 (1995).
- Grallert, B. & Nurse, P. The ORC1 homologue orp1 in fission yeast plays a key role in regulating onset of S phase. *Genes Dev.* **10**, 2644-2654 (1996).
- Masai, H., Miyake, T. & Arai, K. hsk1+, a *Schizosaccharomyces pombe* gene related to *Saccharomyces cerevisiae* CDC7, is required for chromosomal replication. *EMBO J.* **14**, 3094-3104 (1995).
- Fernandez-Sarabia, M., McNerny, C., Harris, P., Gordon, C. & Fantes, P. The cell cycle genes *cdc22* and *suc22* of the fission yeast *Schizosaccharomyces pombe* encode the large and small subunits of ribonucleotide reductase. *Mol. Gen. Genet.* **238**, 241-251 (1993).
- Greenwood, E., Nishitani, H. & Nurse, P. Cdc18p can block mitosis by two independent mechanisms. *J. Cell Sci.* **111**, 3101-3108 (1998).
- Muzi-Falconi, M., Brown, G. W. & Kelly, T. J. *cdc18⁺* regulates initiation of DNA replication in *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA* **93**, 1566-1570 (1996).
- Maierano, D., van Assendelft, G. B. & Kearsey, S. E. Fission yeast *cdc21*, a member of the MCM protein family, is required for onset of S phase and is located in the nucleus throughout the cell cycle. *EMBO J.* **15**, 861-872 (1996).
- Lygerou, Z. & Nurse, P. The fission yeast origin recognition complex is constitutively associated with chromatin and is differentially modified through the cell cycle. *J. Cell Sci.* **112**, 3703-3712 (1999).
- Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795-823 (1991).
- Alfa, C., Fantes, P., Hyams, J., McLeod, M. & Warbrick, E. *Experiments with Fission Yeast—A Laboratory Course Manual* (CSHL Press, New York, 1993).

Acknowledgements

We thank J. Bahler, J. Diffley, D. Griffiths, J. Hayles and H. Murakami for critical reading of the manuscript, T. Orr-Weaver for communicating results before publication, Y. Adachi, T. Carr, S. Kearsey, K. Nasmyth and T. Tanaka for clones and antibodies, and N. Peat for help with figure preparation. This work was supported by the ICRF, an EEC Marie Curie fellowship (Z.L.), and HFSPO and Uehara Foundation fellowships (H.N.).

Correspondence and requests for materials should be addressed to Z.L. (e-mail: z.lygerou@yahoo.com).

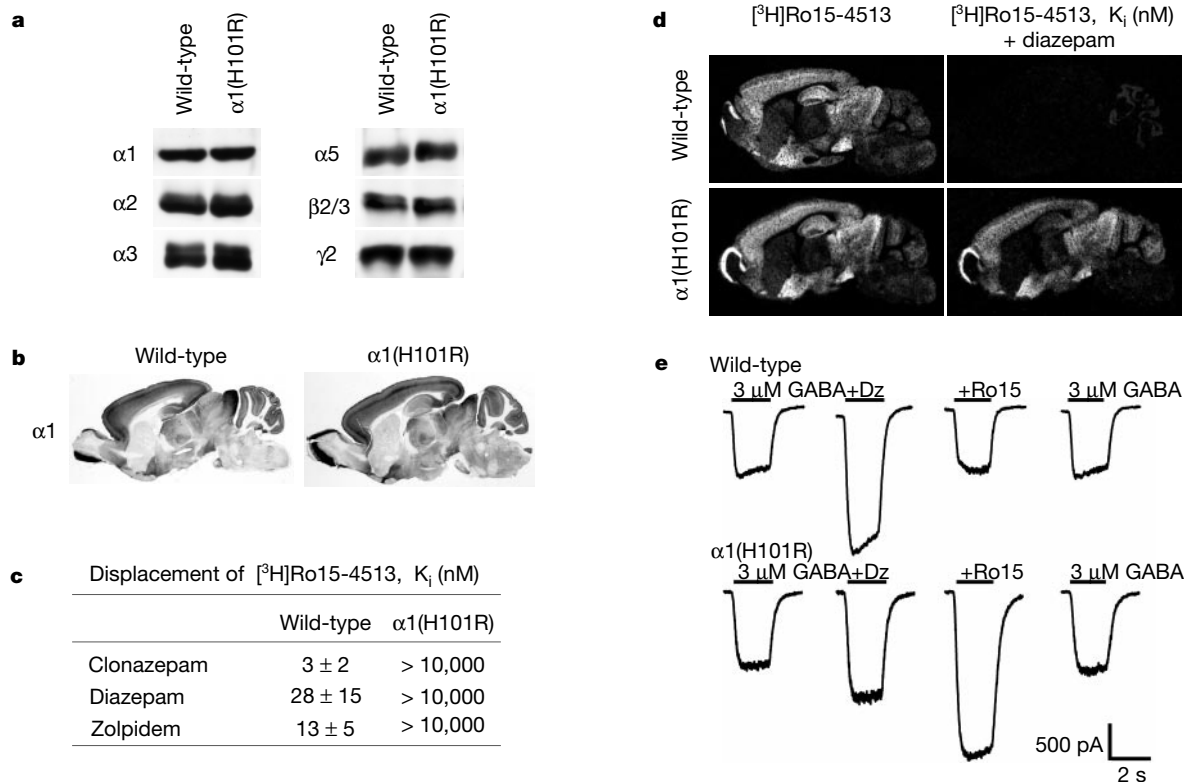
erratum

Benzodiazepine actions mediated by specific γ -aminobutyric acid_A receptor subtypes

Uwe Rudolph, Florence Crestani, Dietmar Benke, Ina Brünig, Jack A. Benson, Jean-Marc Fritschy, James R. Martin, Horst Bluethmann & Hanns Möhler

Nature 401, 796–800 (1999)

The quality of Fig. 2 was unsatisfactory as published. The figure is reproduced again here.



correction

Observation of a square flux-line lattice in the unconventional superconductor Sr₂RuO₄

T. M. Riseman, P. G. Kealey, E. M. Forgan, A. P. Mackenzie, L. M. Galvin, A. W. Tyler, S. L. Lee, C. Ager, D. McK. Paul, C. M. Aegerter, R. Cubitt, Z. Q. Mao, T. Akima & Y. Maeno

Nature 396, 242–245 (1998)

In this Letter, the X-ray Laue patterns used to give the orientation of our crystal were misinterpreted by 45°. With the magnetic field perpendicular to the RuO₂ planes, the nearest-neighbour directions in the square flux-line lattice (FLL) are actually at 45° to the Ru–O–Ru directions in the crystal lattice.

Our values of the superconducting parameters λ and ξ still

suggest that pairing occurs primarily on the electrons on the γ -sheet of the Fermi surface under the conditions of the experiment. However, the orientation of the FLL, interpreted within Agterberg's two-component Ginzburg–Landau (TCGL) theory^{1,2}, is no longer consistent with γ -band pairing if we take the values of Fermi surface anisotropy used in refs 1, 2, and obtained by recent band structure calculations³. However, other calculations (T. Oguchi, private communication) give a different sign of Fermi surface anisotropy. TCGL theory predictions may be modified by extra anisotropy of the superconducting energy gap³, by substantial pairing on the α -and/or β -sheets⁴ or by anisotropy in the electron mass enhancement. Within TCGL theory, the orientation of the FLL that we observe is no longer consistent with the basal plane anisotropy of H_{c2} at low temperatures in a pure sample⁵. H_{c2} is observed to be larger with the field in a {110} direction than a {100}. The nature of the pairing in Sr₂RuO₄ may be more complicated than we suggested previously.

- Agterberg, D. F. *Phys. Rev. Lett.* **80**, 5184–5187 (1998).
- Agterberg, D. F. *Phys. Rev. B* **58**, 14484–14489 (1998).
- Mazin, I. I., Papaconstantopoulos, D. A. & Singh, D. J. Preprint cond-mat/9907442 at (<http://xxx.lanl.gov>) (1999).
- NishiZaki, S., Maeno, Y. & Mao, Z. J. *Low Temp. Phys.* **117**, 1581 (1999).
- Mao, Z. Q., Maeno, Y., Nishizaki, S., Akima, T. & Ishiguro, T. *Phys. Rev. Lett.* **84**, 991–994 (2000).



Cover and side panel illustration

Coloured scanning electron micrograph of fat cells (orange) surrounded and supported by connective tissue fibres (brown) in human adipose connective tissue. (Images courtesy of Quest/SPL.)

We are all familiar with the term 'obesity', but few of us see it as a real disease. The large and manipulative diet industry characterizes obesity as arising from a lack of self-control, and this is a widely held belief. In fact, obesity is a complex disorder of appetite regulation and energy metabolism controlled by specific biological factors. Genes that predispose to obesity in humans and animals have already been identified and indicate the importance of genetic factors in the development of disease.

Five years ago, leptin was identified as the protein responsible for suppressing appetite. It was hailed as a potential wonder drug and catapulted obesity research to the forefront of biomedical science and public imagination. Yet today, in the developed world the incidence of obesity is rising, and there are now as many obese people in the world as there are people suffering from hunger. The financial burden, health risks and impact on quality of life associated with this epidemic warrant a detailed understanding of the molecular mechanisms that regulate body weight, in order to identify new treatments. We therefore devote this collection of reviews — the first in a new, regular feature called 'Nature Insight' — to the molecular biology of obesity.

An overview of the topic is provided by Jeff Friedman on page 632, who with co-workers identified leptin in 1995. (A special news feature on pages 538–540 of this issue examines just how well leptin has lived up to its therapeutic promise.) Obesity has now replaced undernutrition and infectious disease as the most significant contributor to ill health, and Peter Kopelman examines the epidemiology of obesity and its associated medical problems on page 635. The discovery of mutations in human and model organisms, and complementary studies of populations, provide fundamental insight into the primary cause of the disease and are described by Greg Barsh and co-workers on page 644. The thermodynamics of energy expenditure in the mitochondrion and the mechanism of transcriptional control of mitochondrial genes are examined on page 652 by Bradford Lowell and Bruce Spiegelman. Key signalling molecules involved in the hormonal regulation of metabolism by neuronal circuits in the hypothalamus are highlighted by Michael Schwartz and co-workers on page 661, providing insights into energy homeostasis at the molecular level. Finally, when prevention fails, what drugs are available to treat obesity? Although there seems no real cure as yet, on page 672 George Bray and Louis Tartaglia discuss current treatment strategies which aim to inhibit food intake and block fat digestion and provide knowledge of therapeutic agents on the horizon.

We are pleased to acknowledge the financial support of The Roche Group in producing this *Nature Insight*. We share interest and enthusiasm for the advances made in the treatment of obesity. Of course, *Nature* carries the sole responsibility for all editorial content and rigorous peer-review to our normal high standards.

By exploring this disease from its basic epidemiology and associated health risks, through to the molecular mechanisms for regulation and possible therapeutic intervention that may exist, we hope that there is something to be gained by the basic scientist, clinician and the general reader.

Philip Campbell Editor, *Nature*

Ritu Dhand Associate Editor

Obesity in the new millennium

J. M. Friedman

The Rockefeller University, Box 305, 1230 York Avenue, New York, New York 10021-6399, USA

Obesity has increased at an alarming rate in recent years and is now a worldwide public health problem. In addition to suffering poor health and an increased risk of illnesses such as hypertension and heart disease, obese people are often stigmatized socially. But major advances have now been made in identifying the components of the homeostatic system that regulates body weight, including several of the genes responsible for animal and human obesity. A key element of the physiological system is the hormone leptin, which acts on nerve cells in the brain (and elsewhere) to regulate food intake and body weight. The identification of additional molecules that comprise this homeostatic system will provide further insights into the molecular basis of obesity, and possibilities for new treatments.

We live in an era in which advances in medical research reverberate almost instantly through society and culture. This statement is well illustrated by the visibility of recent progress in obesity research. As described in this timely issue, substantial advances have been made towards identifying the components of a physiological system that regulates body weight. Research in this area is at the centre of several important medical and societal issues. First, obesity is a pressing, some consider it the most pressing, health problem in Western and developing countries (see review by Kopelman, pp. 635–643, and ref. 1). Second, family studies of obesity provide a general opportunity for exploring the respective roles of genes and environment in determining human characteristics (see review by Barsh *et al.*, pp. 644–651, and ref. 2). Third, research in this area

has implications for the ways in which alterations of nutritional state affect the function of other organ systems³. Finally, obesity research intersects with considerations of the molecular basis of human behaviour and even the nature of free will (see review by Schwartz *et al.*, pp. 661–671, and refs 3, 4).

Dichotomous views on the causes of obesity

Because eating is an activity in which we all partake, it is not surprising that almost everyone has an opinion about this subject. (To paraphrase Paul Krugman in a recent *New York Times* editorial, this establishes obesity as a political issue.) Two essentially polar explanations characterize most views on the causes of human obesity.

One view suggests that obesity is the result of a fundamental lack of discipline on the part of affected individuals. This view is undoubtedly advanced by the diet industry

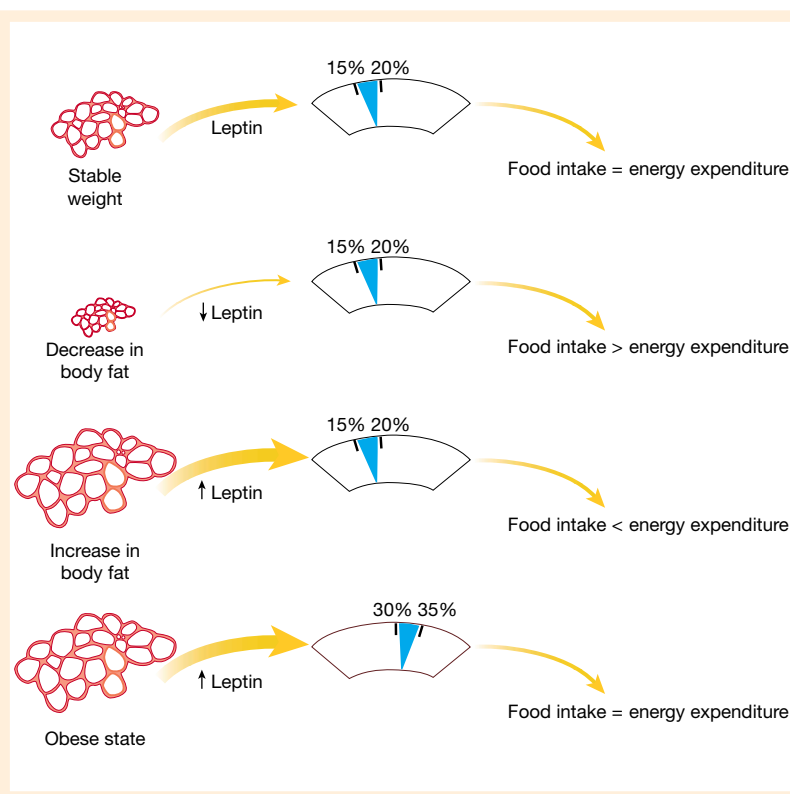


Figure 1 Leptin and the regulation of adipose tissue mass. The cloning of the *ob* gene and the characterization of leptin has indicated that body fat content is under homeostatic control. The available data suggest that leptin is the afferent signal in a feedback loop regulating adipose tissue mass. At an individual's stable weight (shown as 15–20% body fat in this figure, which is the typical fat content of a non-obese subject) the amount of circulating leptin elicits a state in which food intake equals energy expenditure. Increasing leptin levels result in negative energy balance (energy expenditure < food intake), whereas decreasing levels lead to positive energy balance (food intake > energy expenditure). These effects maintain constancy of fat cell mass within a relatively narrow range. Evidence further suggests that the intrinsic sensitivity to leptin is reduced among the obese and that the set point for body fat content is thus increased (designated as 30–35% in the bottom panel). Most obese individuals have high leptin levels and thus enter a state of negative energy balance when weight is reduced and leptin levels fall.

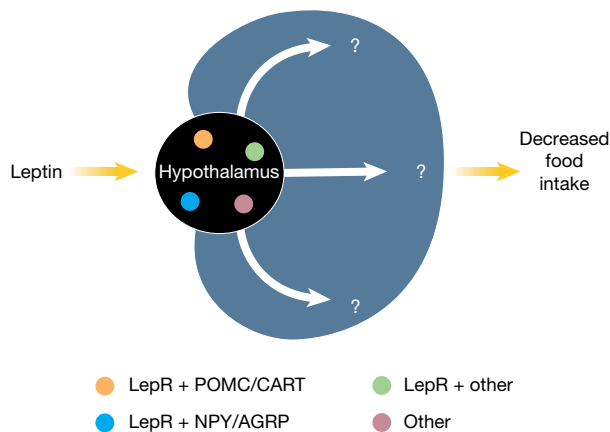


Figure 2 The neural circuit activated by leptin. In the arcuate nucleus of the hypothalamus the leptin receptor is expressed in at least two different classes of neurons. One class expresses NPY and AGRP, two neuropeptides that increase food intake. Another class expresses POMC, the precursor of α -MSH, and CART. Both CART and α -MSH decrease food intake. Other leptin receptor-positive neurons do not express any of these molecules. The available evidence indicates that leptin suppresses the activity of NPY/AGRP neurons and stimulates the activity of POMC/CART neurons. Thus in the absence of leptin the NPY/AGRP neurons are maximally active and food intake is stimulated. In the presence of increased leptin levels the POMC/CART neurons are maximally active and food intake is reduced. When an individual is at their stable weight the activity of these pathways is balanced. The neural mechanisms by which these neurons ultimately change food intake are not known, as represented by the question marks.

which has a financial interest (in aggregate greater than US\$50 billion each year) in promoting the notion that the only thing separating an individual from his or her 'dream' physique is the implementation of a few simple nostrums (provided by the company or author for a modest fee). Although it is true that reducing weight does improve the health of obese and overweight individuals, such remedies fail in the vast majority of cases⁵. Thus more than 90% of individuals who lose weight by dieting eventually return to their original weight.

The alternative view suggests that body weight (or more precisely, the amount of body fat) is physiologically controlled and that deviations in weight in either direction elicit a potent counter-response that resists that change⁶. Implicit in this view is the notion that biological factors determine each individual's body mass, be they lean or obese, and that this state is then defended. The effectiveness of this homeostatic system can be illustrated by a few simple calculations. Over the course of a decade, the weight of an average adult tends to increase slightly. Approximately 10 million kilocalories are consumed over this time. To account for the modest change of weight that is generally observed (assuming the excess weight is deposited as adipose tissue), food intake must precisely match energy output within 0.17% per decade⁷. This extraordinary level of precision has suggested that a robust biological system balances energy input (food intake) and energy expenditure. In recent years, this hypothesis has received substantial experimental support. These data are artfully reviewed in this issue and can be summarized as follows.

A threat to life expectancy

Obesity is formally defined as a significant increase above ideal weight, ideal weight being defined as that which maximizes life expectancy. Actuarial tables indicate that life expectancy is reduced when body-mass index (BMI; defined as mass in kilograms divided by the square of the height in metres), an indicator of adiposity or fatness, is significantly increased above the ideal level (see review by Kopelman, pp. 635–643, and refs 8, 9). This definition formally qualifies 20% of the US population, and a slightly lower percentage of the European population, as obese. There has also been an alarming increase in adolescent obesity in recent years¹⁰. Thus obesity is associated with a significant increase in morbidity and mortality and is a major public health problem. For reasons that are not fully known, obesity is associated with an increased risk of hypertension, heart disease, diabetes and cancer (see review by Kopelman, pp. 635–643). Even modest weight loss ameliorates these associated conditions.

In addition to the prospect of diminished health, obese people are often stigmatized both socially and in the workplace. Although the premium on leanness has become especially prominent in late-twentieth-century Western societies (at least among the affluent), this view is very dependent on the cultural context. In many cultures,

obesity is considered to be a sign of affluence and prestige, particularly among those cultures where food is less available. In modern times, however, intense pressure to be thin is felt by most individuals, lean and obese. Despite this, obesity affects a significant and increasing number of individuals⁸.

Thus the critical question is, within a relatively homogenous environment, why are some individuals lean and others obese? The answer to this question has been informed by the identification of a number of genes that are responsible for animal and human obesity. Twin studies, analyses of familial aggregation and adoption studies all indicate that obesity is largely the result of genetic factors (see review by Barsh *et al.*, pp. 644–651, and ref. 2). Indeed, the heritability of obesity is roughly equivalent to that of height and exceeds that of many disorders that are generally considered to have a genetic basis. The identity of several of these genes is now known and in these instances the evidence that obesity is not simply a personal failing is overwhelming.

Leptin and body-weight regulation

In general, obesity genes encode the molecular components of the physiological system that regulates body weight. A key element of this system is the hormone leptin³. Leptin is produced by fat tissue and reports nutritional information to key regulatory centres in a brain region known as the hypothalamus (Fig. 1). A decrease in body fat leads to a decreased level of this hormone, which in turn stimulates food intake. In addition, decreased leptin levels activate a hormonal response that is characteristic of the starved state¹¹. Increased body fat is associated with increased levels of leptin, which act to reduce food intake. By such a mechanism, weight is maintained within a relatively narrow range. As would be predicted, mutations that result in leptin deficiency are associated with massive obesity in humans as well as rodents^{12,13}. Leptin can also affect energy expenditure, which, in other contexts, is regulated independently of food intake (see review by Lowell and Spiegelman, pp. 652–660, and ref. 14). Changes in leptin concentration have effects on many other organ systems, including reproduction, the immune system and bone formation, which indicates that leptin is an important means by which changes in nutritional state affect physiology³.

Why then are some individuals obese and others not? It seems that the intrinsic sensitivity to leptin is variable and that, in general, obese individuals are leptin-resistant^{3,14,15}. A smaller subset of individuals seems to produce too little leptin. In principle, genetic, environmental and even psychological factors could influence leptin sensitivity or leptin production. The molecular basis for leptin resistance has been explained in some instances. Leptin acts on nerve cells in the brain and modulates their function (Fig. 2). Several key molecules in this neural network are brain peptides known as neuropeptide Y (NPY)

and agouti-related protein (AGRP), which stimulate food intake, and α -melanocyte-stimulating hormone (α -MSH) and cocaine- and amphetamine-regulated transcript (CART), which decrease food intake (see review by Schwartz *et al.*, pp. 661–671, and refs 2, 16, 17). These neural circuits also regulate energy expenditure by means of effects on several key molecules that have recently been identified (see review by Lowell and Spiegelman, pp. 652–660). These effectors include uncoupling proteins and peroxisome proliferator-activated receptor- γ (PPAR- γ) co-activator-1 (PGC-1), a key regulator of the genes that control thermogenesis¹⁸. Genetic evidence indicates that leptin regulates energy balance by modulating the balance among the aforementioned (and other) neuropeptides⁴. Mutations in pro-opiomelanocortin (POMC), the precursor of α -MSH, are associated with obesity¹⁹. In ~3–5% of extremely obese individuals, mutations in an MSH receptor (MC4R) result in a defect in MSH signalling, which causes leptin resistance^{18,20,21}. Mutation in the leptin receptor is also associated with extreme obesity²². In other cases it has been suggested that defective transport of leptin across the blood–brain barrier is the cause of leptin resistance and obesity^{3,23}. Several other factors undoubtedly influence the function of this neural network, and the identification of additional molecules that comprise the neural system will shed more light on the molecular basis of obesity and leptin resistance.

Impact of environmental factors

There is plasticity of this system and such factors as diet, environment, age and perhaps exercise are also important in the pathogenesis of obesity (see review by Kopelman, pp. 635–643, and ref. 24). Thus the system that regulates weight sets a range of body weight in an individual and that range can be further influenced by other factors. For example, it is often noted that the incidence of obesity is rising dramatically in newly Westernized (so-called ‘Coca Cola-nized’) societies. In addition, the advent of a high-fat, Western diet is associated with a staggering increase in body weight among a number of native populations^{24,25}. It is worth noting that a similar trend has also been observed for height in the twentieth century. The average US Civil War soldier was 5 feet 4 inches (1.63 m) tall, yet most people accept that biological factors contribute to differences in stature. The mechanisms by which environmental factors modulate the physiological system that controls weight are poorly understood, but in time they should prove tractable. Environmental factors have been shown to affect leptin sensitivity, as a high-fat diet leads to leptin resistance, although the basis for this is poorly understood³.

What then determines when we eat and how much we eat? Feeding behaviour is complex and is dependent on many factors, including olfactory, visual, emotional and higher cognitive inputs as well as leptin and several other nutritional signals. As the decision of whether or not to eat is the result of neurochemical events in the brain, and not metaphysical, there must be integratory centre(s) that balance all of these inputs. A fuller understanding of these neural events is likely to reveal the mechanisms by which complex behavioural decisions are made, not only for eating, but also perhaps for other complex behaviours.

Tremendous scientific opportunities abound. Recent insights into the molecular mechanisms that regulate weight have already led to numerous possibilities for new treatments and this trend will undoubtedly accelerate (see review by Bray and Tartaglia, pp. 672–677). But for the moment there is no panacea. So the final message is this...weight loss and exercise improve health and should be encouraged. However, a robust biological system makes it exceedingly difficult for most individuals to maintain weight loss for an extended period of time. This fact has deep implications for our perception of obesity and the obese. □

1. Calle, E., Thun, M., Petrelli, J., Rodriguez, C. & Heath, C. Body-mass index and mortality in a prospective cohort of U.S. adults. *N. Engl. J. Med.* **341**, 1097–1105 (1999).
2. Stunkard, A. J., Harris, J. R., Pedersen, N. L. & McClearn, G. E. The body-mass index of twins who have been reared apart. *N. Engl. J. Med.* **322**, 1483–1487 (1990).
3. Friedman, J. M. & Halaas, J. L. Leptin and the regulation of body weight in mammals. *Nature* **395**, 763–770 (1998).
4. Elias, C. *et al.* Leptin differentially regulates NPY and POMC neurons projecting to the lateral hypothalamic area. *Neuron* **23**, 775–786 (1999).
5. Wadden, T. A. Treatment of obesity by moderate and severe caloric restriction. Results of clinical research trials. *Ann. Intern. Med.* **119**, 688–693 (1993).
6. Kennedy, G. C. The role of depot fat in the hypothalamic control of food intake in the rat. *Proc. R. Soc. Lond. B* **140**, 578–592 (1953).
7. Weigle, D. S. Appetite and the regulation of body composition. *FASEB J.* **8**, 302–310 (1994).
8. Kuczmarski, R., Flegal, K., Campbell, S. & Johnson, C. Increasing prevalence of overweight among US adults. *J. Am. Med. Assoc.* **272**, 205–211 (1994).
9. Manson, J. E. *et al.* Body weight and mortality among women. *N. Engl. J. Med.* **333**, 677–685 (1995).
10. Troiano, R. & Flegal, K. Overweight children and adolescents: description, epidemiology, and demographics. *Pediatrics* **101**, 497–504 (1998).
11. Ahima, R. S. *et al.* Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252 (1996).
12. Zhang, Y. *et al.* Positional cloning of the mouse *obese* gene and its human homologue. *Nature* **372**, 425–432 (1994).
13. Montague, C. T. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* **387**, 903–908 (1997).
14. Pellemounter, M. A. *et al.* Effects of the *obese* gene product on body weight regulation in *ob/ob* mice. *Science* **269**, 540–543 (1995).
15. Maffei, M. *et al.* Leptin levels in human and rodent: measurement of plasma leptin and *ob* RNA in obese and weight-reduced subjects. *Nature Med.* **1**, 1155–1161 (1995).
16. Erickson, J. C., Hollopeter, G. & Palmiter, R. D. Attenuation of the obesity syndrome of *ob/ob* mice by the loss of neuropeptide Y. *Science* **274**, 1704–1707 (1996).
17. Fan, W., Boston, B. A., Kesterson, R. A., Hruby, V. J. & Cone, R. D. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* **385**, 165–168 (1997).
18. Puigserver, P. *et al.* A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* **92**, 829–839 (1998).
19. Krude, H. *et al.* Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nature Genet.* **19**, 155 (1998).
20. Yeo, G. *et al.* A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nature Genet.* **20**, 111–112 (1998).
21. Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nature Genet.* **20**, 113–114 (1998).
22. Clement, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398–401 (1998).
23. Schwartz, M. W., Peskind, E., Raskind, M., Boyko, E. J. & Porte, D. Jr Cerebrospinal fluid leptin levels: relationship to plasma levels and to adiposity in humans. *Nature Med.* **2**, 589–593 (1996).
24. James, W. in *The Origins and Consequences of Obesity* (eds Chadwick, D. & Cardew, G.) 1–16 (Wiley, Chichester, 1996).
25. Ravussin, E. Metabolic differences and the development of obesity. *Metabolism* **9**, 12–14 (1995).

Acknowledgements

I thank S. Korres for assistance in preparing this manuscript and J. Breslow, J. Darnell, S. Friedman, A. Soukas, M. Stoffel and E. Ruppel-Shell Stoffel for critical and thoughtful comments.

Obesity as a medical problem

Peter G. Kopelman

*St Bartholomew's & The Royal London School of Medicine, Queen Mary & Westfield College, London E1 2AD, UK
(e-mail: P.G.Kopelman@mds.qmw.ac.uk)*

Obesity is now so common within the world's population that it is beginning to replace undernutrition and infectious diseases as the most significant contributor to ill health. In particular, obesity is associated with diabetes mellitus, coronary heart disease, certain forms of cancer, and sleep-breathing disorders. Obesity is defined by a body-mass index (weight divided by square of the height) of 30 kg m^{-2} or greater, but this does not take into account the morbidity and mortality associated with more modest degrees of overweight, nor the detrimental effect of intra-abdominal fat. The global epidemic of obesity results from a combination of genetic susceptibility, increased availability of high-energy foods and decreased requirement for physical activity in modern society. Obesity should no longer be regarded simply as a cosmetic problem affecting certain individuals, but an epidemic that threatens global well being.

Obesity causes or exacerbates many health problems, both independently and in association with other diseases¹. In particular, it is associated with the development of type 2 diabetes mellitus, coronary heart disease (CHD), an increased incidence of certain forms of cancer, respiratory complications (obstructive sleep apnoea) and osteoarthritis of large and small joints. The Build and Blood Pressure Study has shown that the adverse effects of excess weight tend to be delayed, sometimes for ten years or longer². Life-insurance data and epidemiological studies confirm that increasing degrees of overweight and obesity are important predictors of decreased longevity³. In the Framingham Heart Study, the risk of death within 26 years increased by 1% for each extra pound (0.45 kg) increase in weight between the ages of 30 years and 42 years, and by 2% between the ages of 50 years and 62 years⁴. Despite this evidence, many clinicians consider obesity to be a self-inflicted condition of little medical significance. Here I will review the epidemiology and factors influencing obesity and the health consequences of excessive body fat.

Definition of overweight and obesity

In clinical practice, body fat is most commonly and simply estimated by using a formula that combines weight and height. The underlying assumption is that most variation in weight for persons of the same height is due to fat mass, and

the formula most frequently used in epidemiological studies is body-mass index (BMI). Box 1 details the practical methods used in clinical practice to assess body fatness. A graded classification of overweight and obesity using BMI values provides valuable information about increasing body fatness. It allows meaningful comparisons of weight status within and between populations and the identification of individuals and groups at risk of morbidity and mortality. It also permits identification of priorities for intervention at an individual or community level and for evaluating the effectiveness of such interventions. It is important to appreciate that, owing to differences in body proportions, BMI may not correspond to the same degree of fatness across different populations. Nor does it account for the wide variation in the nature of obesity between different individuals and populations. A World Health Organization (WHO) expert committee has proposed the classification of overweight and obesity that applies to both men and women and to all adult age groups (Table 1)^{5,6}.

Defining a 'healthy weight' for a particular society presents problems. First, the definition is based on total mortality rates, which can be misleading. People frequently lose weight as a consequence of illness, which may go unrecognized at the time of survey, but results in death. This implies a higher mortality among those with lower weights and is referred to as reverse causation. A second major concern is the confounding factors, such as smoking, that

Box 1

Defining obesity

Method	Definition	Advantages/limitations
BMI	Weight in kilograms divided by square of the height in metres	BMI correlated strongly with densitometry measurements of fat mass; main limitation is that it does not distinguish fat mass from lean mass
Waist circumference	Measured (in centimetres) at midpoint between lower border of ribs and upper border of the pelvis	Waist circumference and waist-to-hip ratio provide measures for assessing upper body fat deposition; neither provide precise estimates of intra-abdominal (visceral) fat
Skinfold thickness	Measurement of skinfold thickness (in centimetres) with callipers provides a more precise assessment if taken at multiple sites	Measurements are subject to considerable variation between observers, require accurate callipers and do not provide any information on abdominal and intramuscular fat
Bioimpedance	Based on the principle that lean mass conducts current better than fat mass because it is primarily an electrolyte solution; measurement of resistance to a weak current (impedance) applied across extremities provides an estimate of body fat using an empirically derived equation	Devices are simple and practical but neither measure fat nor predict biological outcomes more accurately than simpler anthropometric measurements

Table 1 Cut-off points proposed by a WHO expert committee for the classification of overweight

BMI* (kg m ⁻²)	WHO classification	Popular description
<18.5	Underweight	Thin
18.5–24.9	—	'Healthy', 'normal', 'acceptable'
25.0–29.9	Grade 1 overweight	Overweight
30.0–39.9	Grade 2 overweight	Obesity
≥40.0	Grade 3 overweight	Morbid obesity

*BMI is the weight in kilograms divided by the square of the height in metres.

The data presented in Tables 1 and 2 reflect knowledge acquired largely from epidemiological studies in developed countries. Preliminary information from developing nations indicates that lower cut-off levels for both BMI and waist circumference (see Table 2) are necessary for certain populations who are at particular risk from comparatively modest degrees of overweight.

may distort the association between body weight and mortality. The Nurses Health Study, which prospectively studied 116,000 women in the United States during a 17-year period, shows a U-shaped relationship between mortality and BMI in an overall age-adjusted analysis. However, the relationship becomes a simple positive association when reverse causation is accounted for and the analysis limited to those who had never smoked⁷.

Despite these shortcomings in the calculation, there is a close relationship between BMI and the incidence of several chronic conditions caused by excess fat (Fig. 1), including type 2 diabetes, hypertension, CHD and cholelithiasis. This relationship is approximately linear for a range of BMI indexes less than 30 (kg m⁻²), but all risks are greatly increased for those subjects with a BMI above 29, independent of gender^{8,9}.

Waist circumference correlates with measures of risk for CHD such as hypertension or blood lipid levels. The choice of cut-off points on the waist circumference continuum involves a trade-off between sensitivity and specificity similar to that for BMI. Gender-specific cut-off points for waist circumference may be of guidance in interpreting values for adults: proposed cut-off levels are shown in Table 2, with level 1 being intended to alert clinicians to potential risk, whereas level 2 should initiate therapeutic action¹⁰.

Epidemiology of overweight and obesity

Obesity can be defined as a disease in which excess body fat has accumulated such that health may be adversely affected. Conservative estimates of the economic costs of obesity in developed countries are between 2 and 7% of the total health costs, which represents a significant expenditure of national health-care budgets¹¹. It is highly beneficial to be able to estimate prevalence and secular trends in obesity in order to identify those at risk and assist policy makers and public-health planners. The major health consequences of obesity are predictable from an understanding of the pathophysiology of increasing body fat. Obese individuals with excess fat in intra-abdominal depots are at particular risk of negative health consequences, with certain ethnic populations carrying different levels of risk¹². To make true comparisons of the burden of obesity between countries it is necessary to compare population-based data on measured height and weight that followed identical protocols for measurement and collection during the same time period.

The range of BMI of a population varies significantly according to the stage of economic transition and associated industrialization of a

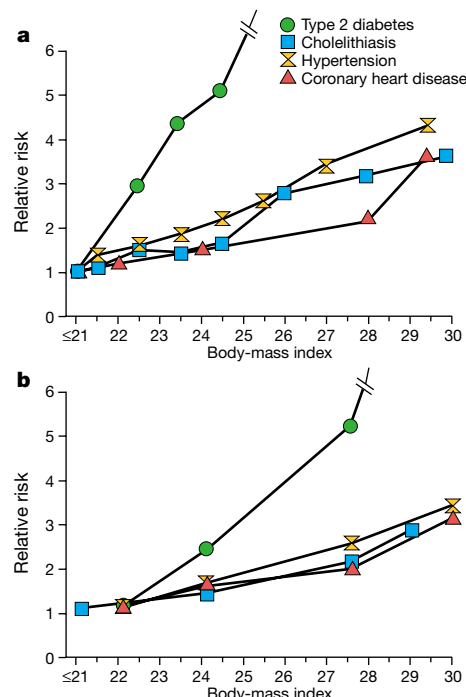


Figure 1 Relation between BMI up to 30 and the relative risk of type 2 diabetes, hypertension, CHD and cholelithiasis. **a**, Relations for women, initially 30 to 55 years old, who were followed up for 18 years. **b**, Relations for men, initially 40 to 65 years old, who were followed up to ten years. Courtesy of the authors in ref. 9, and the editor, *New England Journal of Medicine*.

country (such as a shift from dietary deficit to one of dietary excess). As the proportion of the population with a low BMI decreases there is an almost symmetrical increase in the population with a BMI above 25. This indicates the tendency for a population-wide shift as socio-economic conditions improve, with overweight replacing thinness. In the first stages of the transition, wealthier sections of society show an increase in the proportion of people with a high BMI, whereas thinness remains the main concern among the less wealthy. The distribution of BMI tends to change again in the later phases of transition with an increasing prevalence of high BMI among the poor. Importantly, changes in adult prevalence of obesity are reflected by a striking increase in childhood and adolescent weight in both industrialized and developing countries. The early onset of obesity leads to an increased likelihood of obesity in later life as well as an increased prevalence of obesity-related disorders^{13,14}.

Obesity (defined as a BMI above 30) is a common condition in every continent (Fig. 2). The most comprehensive information in Europe derives from data collected between 1983 and 1986 for the MONICA study¹⁵. On average, 15% of men and 22% of women were obese, with overweight also being more common among women than men. More than half the adult population between 35 and 65 years of age in Europe were either overweight or obese. In England and Wales the most recent health survey has confirmed an increase in the prevalence of obesity in adults from 6% in men and 8% in women in 1980 to 17% of men and 20% of women in 1997¹⁶. National surveys in the United States have shown a marked increase in prevalence of obesity over time. The striking increase in prevalence between 1980 and 1994 confirms that population-wide increases in overweight and obesity may occur over a short period of time. The most recent data from the United States, derived from the third National Health and Nutrition Examination Survey (1988–94), shows ~20% of US men and ~25% of US women are obese¹⁷. Detailed sub-analysis shows African-American women and other minority populations to be

Table 2 Waist circumference predicts risk of metabolic complications

	Increased risk	Substantially increased risk
Men	≥94 cm	≥102 cm
Women	≥88 cm	≥88 cm

Gender-specific waist circumferences are presented that denote 'increased risk' (level 1) and 'substantially increased risk' (level 2) of metabolic complications associated with obesity in Caucasians. Level 1 is intended to alert clinicians to potential risk for CHD whereas level 2 should initiate therapeutic action.

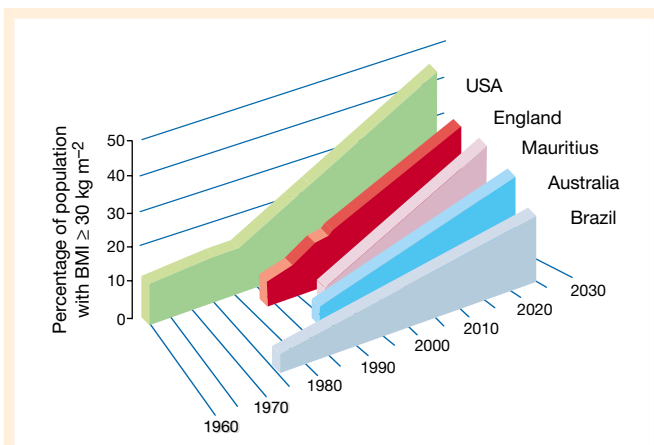


Figure 2 Historic, current and projected obesity prevalence rates ($\text{BMI} \geq 30 \text{ kg m}^{-2}$) for the United States, England and Wales, Mauritius, Australia and Brazil from 1960 to 2025. Courtesy of the International Obesity Task Force.

particularly susceptible. Obesity is also prevalent in Latin America and a particular problem in the Caribbean¹⁸.

But the increasing prevalence of obesity is not confined merely to Europe and the Americas. In Southeast Asia a marked rise is being seen in all populations, and in Japan and China a pronounced increase in the prevalence of overweight and obesity has been observed during the past two decades¹⁹. Obesity is now more prevalent in Malaysia than undernutrition in both urban and rural communities, but the most striking figures come from the Pacific region. In urban Samoa the prevalence of obesity is estimated as greater than 75% of adult women and 60% of adult men²⁰. High prevalence rates also occur in the Middle East. In the United Arab Emirates obesity is recognized as a major public-health problem that may be important in the increasing occurrence of other chronic diseases²¹.

Factors influencing obesity

Obesity is not a single disorder but a heterogeneous group of conditions with multiple causes. Body weight is determined by an interaction between genetic, environmental and psychosocial factors acting through the physiological mediators of energy intake and expenditure. Although genetic differences are of undoubted importance, the marked rise in the prevalence of obesity is best explained by behavioural and environmental changes that have resulted from technological advances (Fig. 3).

Genetics

Fatness runs in families but the influence of the genotype on the aetiology of obesity may be attenuated or exacerbated by nongenetic factors. Apart from rare obesity-associated syndromes, the genetic influences seem to operate through susceptibility genes. Such genes increase the risk of developing a characteristic but are not essential for its expression or, by themselves, sufficient to explain the development of a disease. The susceptible-gene hypothesis is supported by findings from twin studies in which pairs of twins were exposed to periods of positive and negative energy balance²². The differences in the rate of weight gain, the proportion of weight gained and the site of fat deposition showed greater similarity within pairs than between pairs. This suggests differences in genetic susceptibility within a population determine those who are most likely to become obese in any given set of environmental circumstances.

A candidate gene is defined as that part of the DNA molecule that directs the synthesis of a specific polypeptide chain closely associated with a particular disease. The search for obesity genes requires a multifaceted approach that involves studies of potential candidate genes derived from animal models, human obesity syndromes and a genome-wide search using microsatellites covering the human genome. Candidate genes for obesity can be chosen for their possible

effects on body fat composition, anatomical distribution of fat, food intake and energy expenditure.

Monogenic rodent models of obesity are all characterized by early onset of obesity, hyperinsulinaemia and insulin resistance. The genetic aetiology of obesity in the laboratory-bred *ob* mouse is well defined. The *ob* gene is positioned on chromosome 6 and expressed exclusively in adipose tissue in normal mice. The gene product, which is called leptin (derived from Greek *leptos*, meaning thin), is nonfunctional in mice that are homozygous for the *ob* mutation²³. Replacement of leptin by intraperitoneal injections in these animals leads to a reduction in body weight, body fat, food intake and serum insulin. Leptin introduced into the lateral or third ventricle of the brain is effective in reducing weight, indicating a probable central effect²⁴. By contrast, the administration of leptin to the *db/db* mouse (an obese mice characterized by high leptin levels) has no effect on appetite, body weight or body fat. These mice have a mutation of the leptin-receptor gene, which gives rise to a nonfunctioning leptin receptor. The initial hypothesis that obesity in humans results from a relative or absolute deficiency of leptin has not been borne out. Paradoxically, most obese humans have high circulating levels of leptin that are raised in proportion to fat mass²⁵, whereas only a handful of individuals with severe obesity have been identified either with congenital deficiency or a mutation in the leptin-receptor gene. Detailed observations of a child with severe, early-onset obesity treated with subcutaneous injections of leptin has demonstrated significant and impressive weight loss without any alteration in energy expenditure over a 24-hour period — a reduction in basal metabolic rate was counterbalanced by an increase in physical activity²⁶. The main effect of leptin in inducing weight loss was mediated by its suppressive effect on food intake. These findings raise important questions about the primary role of leptin in humans and demonstrate the complexity of human hypothalamic function compared to rodents.

Several candidate genes have been associated with human obesity or its metabolic complications. They include receptors that are important in mechanisms of thermogenesis (for example, β_3 -adren-ergic-receptor gene and the family of uncoupling proteins) as well as those involved in appetite regulation.

Obesity is a consistent finding in many single-gene disorders of humans. One example is Prader–Willi syndrome (PWS), which is characterized by upper-body obesity, short stature, mental retardation and hypogonadism. The condition is seen in approximately 1 in 25,000 births, may occur ‘sporadically’, but usually is associated with a familial inheritance: it is caused by a deletion of the paternal segment of chromosome 15. A second, less common disorder is Bardet–Biedl syndrome. Although sharing many characteristics with PWS, studies of affected families have identified several different chromosomal loci (chromosomes 16, 11, 3 and 15) responsible for the syndrome, which confirms the heterogeneity of the condition.

The advent of marker libraries covering the entire human genome is providing an opportunity for random genome-wide search for candidate genes contributing to human obesity through the study of large numbers of individuals within defined populations or families. These approaches are considered in the review by Barsh *et al.*, pp. 644–651.

Environmental factors

Implicit to the susceptible-gene hypothesis is the role of environmental factors that unmask latent tendencies to develop obesity. Predictions about possible interactions between genes and the environment are difficult because there may be a delay in an individual's exposure to an ‘obesogenic’ environment, and/or alteration in lifestyle related to living circumstances and uncertainty about the precise timing of the onset of weight gain.

Energy expenditure. The most variable component of energy expenditure is physical activity, representing 20–50% of total energy expenditure. The analysis of the level of physical activity is

similar in groups of subjects with a BMI of <20, 20–25 and 25–35, which indicates similar levels of habitual activity. The measurement of energy expenditure within the home, using doubly labelled water, also shows comparable values between obese and lean subjects when corrected for different body sizes²⁷.

A defect in metabolic mechanisms that control energy expenditure has not been described in human obesity. Longitudinal studies of Pima Indians indicate that the risk of 10-kg weight gain during a 4-year follow-up is sevenfold higher in those in the lowest tertile of relative resting metabolic rate (RMR) compared with those in the highest tertile²⁸. Nevertheless, even in this population, which is predisposed to obesity, this predicts only 40% of the weight gain. No association has been observed between RMR and 10-year weight gain in a Dutch population and results from other studies have also questioned the validity of this putative association²⁹.

Cross-cultural studies of physical activity and BMI demonstrate a sevenfold increased risk of overweight (BMI > 25) in those with a physical activity level ratio (total energy expenditure/RMR) of <1.8 (ref. 30). In developed countries there is a relationship between low levels of physical activity and obesity. A longitudinal Finnish study found that those reporting physical exercise three or more times each week had on average lost weight since a preceding survey. By contrast, those who undertook little physical activity gained weight and had twice the risk of gaining 5 kg or more³¹. In Finland, a decline in physical activity at work and in transport during the past 10 years has been accompanied by a significant increase in leisure time. Among children in the United States, the relative risk of obesity is 5.3 times greater for children who watch television for 5 h or more each day compared with those children who watch for less than 2 h, even after correcting for a wide range of socioeconomic variables³².

In the United Kingdom, a study combining data on energy intake and physical activity in relation to the secular increase in adult obesity shows no relationship between total energy intake or fat consumption and the prevalence of obesity, but a close relationship between proxy measures of physical activity (television viewing and car ownership)³³.

Energy intake. It is surprising that no direct correlation has been reported between the prevalence of obesity and increased energy intake in developed nations, given the ready availability of highly palatable foods. The understanding of the role of energy intake in the aetiology of obesity is confounded by failure to report food intake accurately. Under-reporting is widely recognized as a feature of obesity, with comparisons of energy intake and expenditure in obese subjects showing a consistent shortfall in self-reported food intake of approximately 30% of the energy requirements^{34,35}. There is good evidence that individual macronutrients (protein, fat and carbohydrate) exert differing effects on eating behaviour predominantly as a result of their effects on satiety. Fat has a weak satiating capacity, particularly when compared with protein, and subjects in experimental situations readily overeat when presented with high-fat foods³⁶.

It seems likely that environmental influences act through increasing energy intake and/or decreasing energy expenditure. There is some evidence that high-fat diets are associated with an increased risk of obesity within populations, but cross-cultural dietary studies have failed to show any consistent relationship between nutritional factors and relative weights³⁷.

Culture. Evidence for the critical role of environmental factors in the development of obesity comes from migrant studies and the 'westernization' of diet and lifestyles in developing countries. The pronounced increase in age-standardized prevalence of obesity (>60% in men and women) in the Naurians in Micronesia and Polynesians in Western Samoa is closely paralleled by alterations in diet and lifestyle³⁸. A marked change in BMI is frequently witnessed in migrant studies, where populations with a common genetic heritage live under new and different environmental circumstances. Pima Indians, for example, living in the United States are on average 25 kg heavier than Pima Indians living in Mexico³⁹. A similar trend is

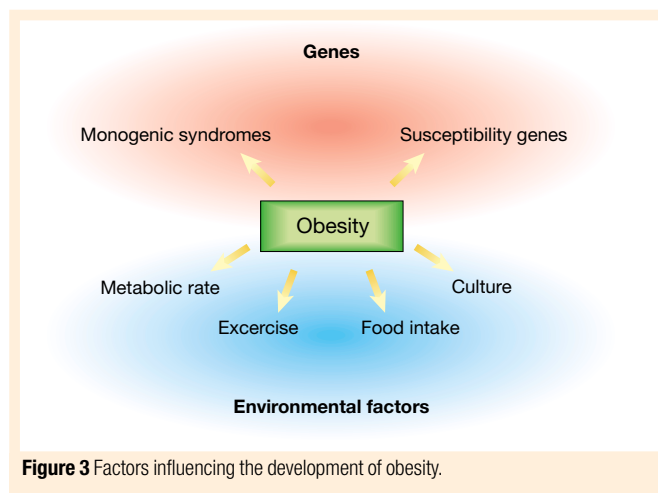


Figure 3 Factors influencing the development of obesity.

seen for Africans living in the United States. In Nigeria the mean BMI for men and women is 21.7 and 22.6, respectively; in the United States the average BMI for Nigerian men and women is 27.1 and 30.8, respectively⁴⁰. This increasing prevalence of obesity is associated with adverse health consequences. The prevalence of hypertension in adult Nigerians living in Africa is 15%, whereas it is as high as 30% among those living in United States.

In both men and women the prevalence of overweight and obesity increases with age until 50 to 60 years; it is particularly apparent between the ages of 20 and 40 years. There are large, usually unexplained variations between ethnic groups — this is particularly apparent in US women with the rapidity of change occurring with increasing affluence of particular lower economic groups (22% of Caucasian women are obese, 30% of African-American women and 34% of Mexican-American women)¹⁷. In industrialized countries, a higher prevalence of overweight and obesity is observed in those with lower educational attainments and low income, although the reverse may be seen in developing countries. There is a tendency for overweight to increase after marriage and with increasing parity. Dietary intake and physical activity are crucially important factors in increasingly affluent societies.

Analysis of the prevalence of obesity by socioeconomic status in England and Wales demonstrates a strong gradient related to social class, especially in women, ranging from 10.7% in social class I (high) to 25% in social class V (low). Interestingly, this is accompanied by marked differences in measures of physical activity with social classes IV and V spending significantly more time watching television and being more likely to define themselves as inactive compared with those in social class I (ref. 41).

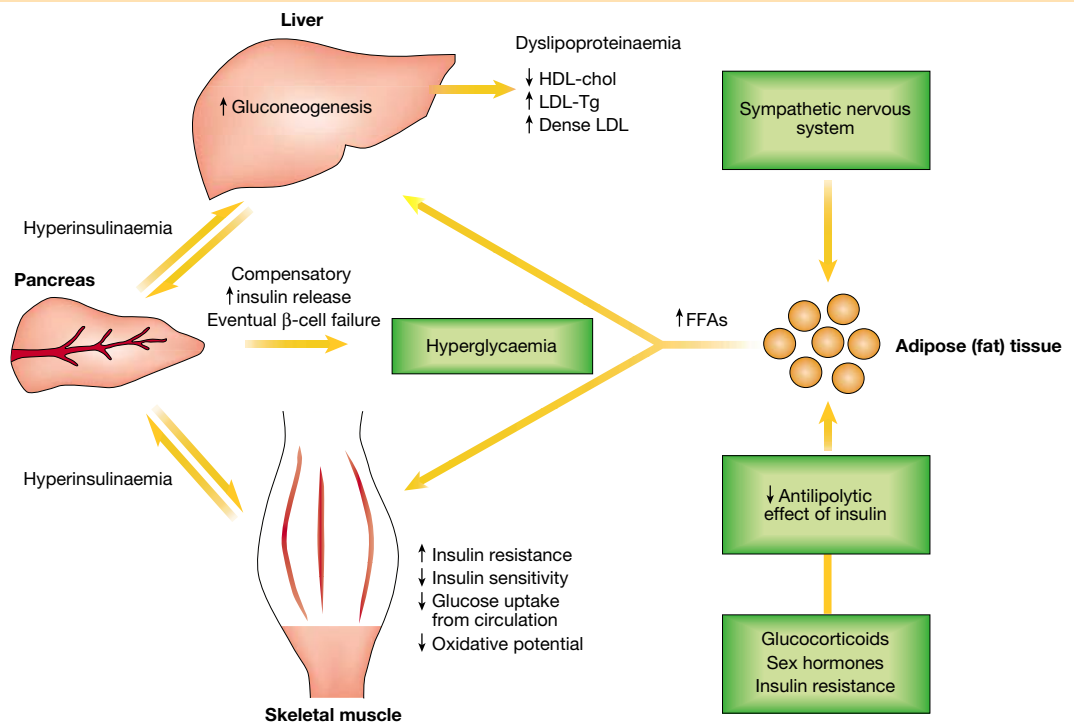
Fetal nutrition. Evidence indicates that undernutrition of the fetus during intrauterine development may determine the later onset of obesity, hypertension and type 2 diabetes independent of genetic inheritance. Such a phenomenon suggests the possibility of long-term programming of genetic expression as a consequence of altered intrauterine growth⁴². Barker has hypothesized that an adverse nutritional environment *in utero* causes defects in the development of body organs leading to a 'programmed' susceptibility that interacts with later diet and environmental stresses to cause overt disease many decades later. In support of the hypothesis is the finding of an inverse relationship between birthweight and systolic blood pressure in both men and women in later life, with the highest mean systolic blood pressures being observed in those with the lowest birthweight and highest current weight⁴³.

Central to the ('thrifty phenotype') hypothesis is the view that a predisposition to type 2 diabetes and other conditions, including adult obesity, is an adaptation to malnutrition by the developing fetus. Low birthweight is a proxy for a variety of intrauterine influences but predominantly is caused by maternal malnutrition. It is

Box 2

Obesity and type 2 diabetes

Schematic illustration of the detrimental effect of increasing amounts of adipose (fat) tissue on whole-body sensitivity to the actions of insulin and glucose tolerance. Elevated rates of fat breakdown (lipolysis) lead to a release of FFAs. These have a detrimental action on the uptake of insulin by the liver, which in turn results in increased gluconeogenesis (breakdown of amino acids and conversion to glucose), production of glucose by the liver, and systemic dyslipidaemia. These factors contribute to the prevailing systemic hyperinsulinaemia (raised circulatory insulin concentrations) and decreased skeletal insulin sensitivity with reduced glucose uptake. Initially, the β -cells of the pancreas compensate for these processes by producing more insulin. In time, there is failure of the β -cells and the development of a raised circulating blood glucose concentration (hyperglycaemia), and hence type 2 diabetes.



Key:

- ↑ **Gluconeogenesis** — increased breakdown of amino acids and conversion to glucose.
- Hyperglycaemia** — raised circulating blood glucose concentration.
- Hyperinsulinaemia** — raised circulating blood insulin concentration.
- ↓ **Insulin resistance**–**insulin sensitivity** — reduced uptake by tissues of circulating glucose due to decreased sensitivity of cells to insulin action.
- ↓ **Antilipolytic effect of insulin** — diminished action of insulin on fat cells to prevent the breakdown of fat to free fatty acids (FFAs).
- Dyslipoproteinaemia** — abnormal blood lipid concentrations: ↓ HDL-cholesterol; ↑ LDL-Tg, increased LDL-triglyceride; ↑ LDL, increased LDL concentrations.

suggested that the fetus adapts its growth and metabolism to the expectation of poor availability of nutrition postnatally. This may have survival advantages *in utero* by targeting available nutrients to essential organs and, in later life, by increasing the ability to store energy as fat to provide energy reserves for use when food is scarce. These adaptations are detrimental when there is a constant supply of nutrition. Studies of adult subjects with documented low birthweights have indicated that they are almost seven times more likely to have either impaired glucose tolerance or type 2 diabetes at age 64 years compared with those born the heaviest⁴⁴. Importantly, the highest blood glucose concentrations are found in men who were lightest at one year of age but who had the highest BMI when aged 64 years. There are a number of mechanisms through which the hypothesis could work. These include a deficiency of insulin production by pancreatic β -cells and alterations in placental vasculature resulting in poor maternal transfer of nutrients. Fetal and early life are critical periods for pancreatic β -cell development because about half the adult mass of β -cells is present by one year of age. In addition, animal studies have shown that maternal protein restriction during pregnancy can markedly reduce pancreatic vascularization in the offspring.

There are reports showing an inverse correlation between abdominal fatness and birthweight but none which have examined the effect of size at birth and the subsequent incidence of obesity. Nevertheless,

the study of cohorts born around the time of the time of the Dutch famine during the winter of 1944–1945 provide some of the most convincing evidence that both early and late gestation are critical periods for the subsequent development of obesity. Compared with a control group not exposed to famine during pregnancy, the prevalence of obesity was significantly higher in those adults whose fetal exposure to famine coincided with the first two trimesters of pregnancy. In contrast, the prevalence of obesity was significantly lower in those whose exposure to famine occurred in the third trimester or shortly after birth⁴⁵. These findings are consistent with an appetite rebound after the famine when mothers were able to eat normally towards the end of pregnancy at the critical time of fetal fat accumulation.

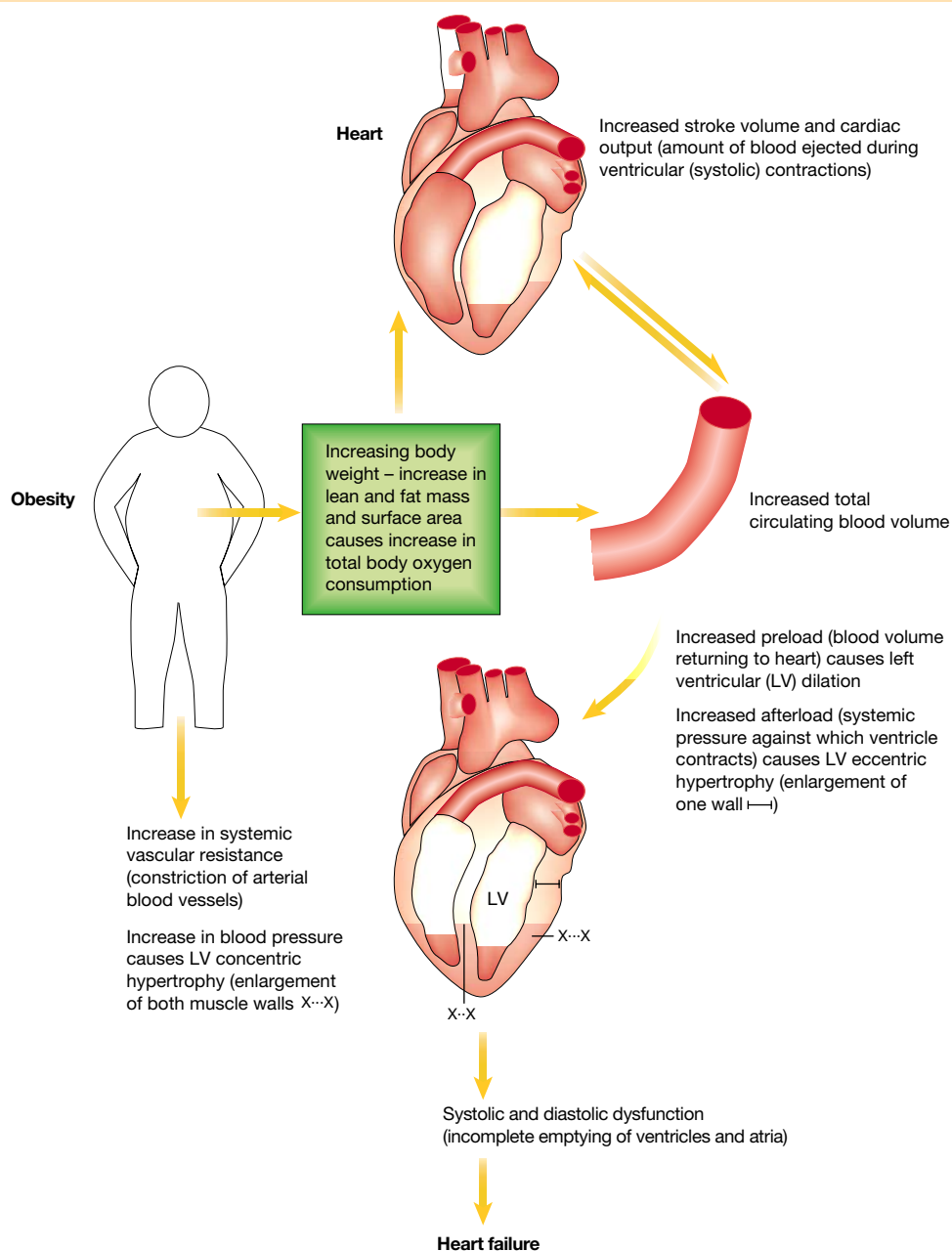
Obesity as a medical problem

Increasing body fatness is accompanied by profound changes in physiological function. These changes are, to a certain extent, dependent on the regional distribution of adipose tissue. Generalized obesity results in alterations in total blood volume and cardiac function, whereas the distribution of fat around the thoracic cage and abdomen restricts respiratory excursion and alters respiratory function. The intra-abdominal visceral deposition of adipose tissue, which characterizes upper body obesity, is a major contributor to the development of hypertension, elevated plasma insulin

Box 3

Cardiovascular function in obesity

Schematic illustration of the physiological changes in cardiovascular function that accompany progressive weight gain. Increasing body weight is associated with an increase in both lean and fat mass and in surface area. This, and the associated increase in total blood volume, is in turn accompanied by an increase in stroke volume and cardiac output. The increase in circulatory preload and afterload lead to left ventricular (LV) dilatation and eccentric hypertrophy. An increase in systemic vascular resistance seen in some obese individuals results in a sustained rise in blood pressure (hypertension) and concentric LV hypertrophy. In time the atria and ventricles begin to fail with the development of heart failure.



concentrations and insulin resistance, diabetes mellitus and hyperlipidaemia.

Obesity and type 2 diabetes mellitus

Obesity is characterized by elevated fasting plasma insulin and an exaggerated insulin response to an oral glucose load⁴⁶. Overall fatness and the distribution of body fat influence glucose metabolism through independent but additive mechanisms. Increasing upper body obesity is accompanied by a progressive increase in the glucose and insulin response to an oral glucose challenge with a positive correlation being observed between increasing upper body obesity and measures of insulin resistance. Post-hepatic insulin delivery is increased in upper body obesity leading to more marked peripheral insulin concentrations that, in turn, lead to peripheral insulin resistance (Box 2).

Different fat depots vary in their responsiveness to hormones that

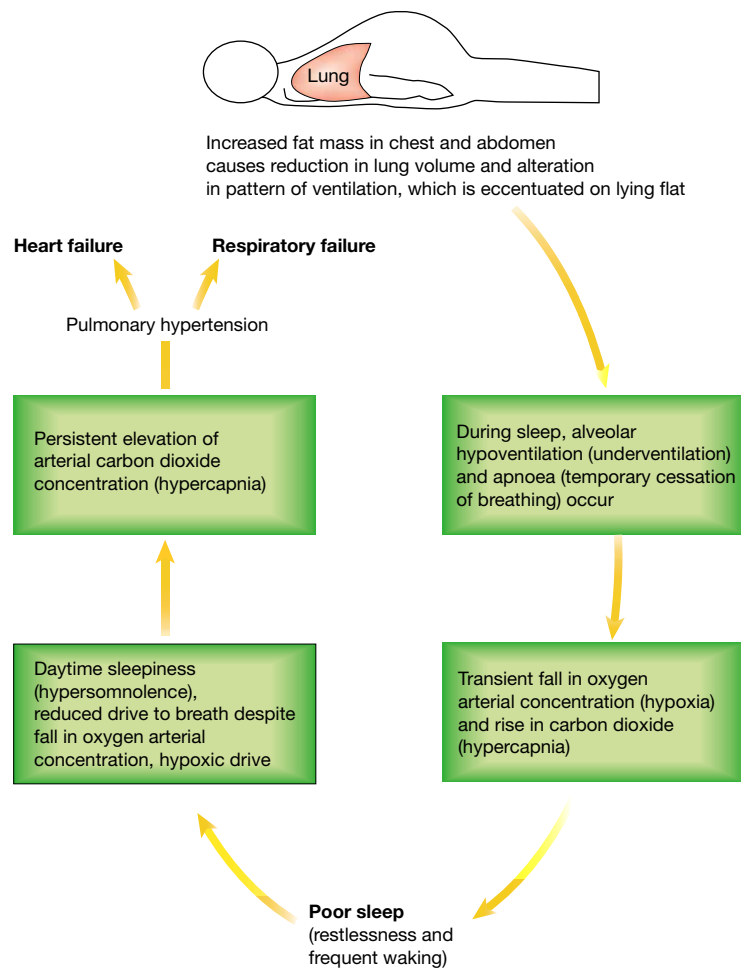
regulate lipolysis and this also varies according to fat distribution⁴⁷. In both men and women, the lipolytic response to noradrenaline is more marked in abdominal than gluteal or femoral adipose tissue⁴⁸. Cortisol may also contribute to this enhanced lipolysis by further inhibiting the antilipolytic effect of insulin. These factors contribute to an exaggerated release of free fatty acids (FFAs) from abdominal adipocytes into the portal system⁴⁹. FFAs have a deleterious effect on insulin uptake by the liver and contribute to the increased hepatic gluconeogenesis and hepatic glucose release observed in upper body obesity. Insulin insensitivity is confined not only to adipocytes — the process being accentuated by insulin resistance of skeletal muscle.

The elevation in plasma FFA concentration, particularly postprandially when they are usually suppressed by insulin, leads to an inappropriate maintenance of glucose production and an impairment of hepatic glucose utilization (impaired glucose tolerance).

Box 4

Sleep-breathing abnormalities in obesity

Schematic illustration of the changes in respiratory function during sleep that result from a progressive increase in body weight. Respiratory function is compromised in obese subjects, especially when they lie flat. During sleep, there is alveolar hypoventilation and transient episodes of apnoea that are accompanied by a fall in arterial oxygen saturation (hypoxia) and a rise in arterial carbon dioxide (hypercapnia). In some individuals, these factors lead to daytime sleepiness (hypersomnolence) with persistent hypoxia and hypercapnia accompanied by the development of pulmonary hypertension, heart failure and, eventually, respiratory failure.



Reduced hepatic clearance of insulin leads to increased peripheral (systemic) insulin concentrations and to a further downregulation of insulin receptors.

In the initial phases of this process, the pancreas can respond by maintaining a state of compensatory hyperinsulinaemia with gross decompensation of glucose tolerance being prevented. With ever increasing plasma concentrations of FFAs, the insulin-resistant individual cannot continue to maintain this state of compensatory hyperinsulinaemia, and hyperglycaemia prevails. Hyperinsulinaemia and insulin resistance are both significant correlates of a dyslipoproteinaemic state and contribute to the characteristic alterations of plasma lipid profile associated with obesity: elevated fasting plasma triglyceride concentration, reduced high-density lipoprotein-cholesterol, marginal elevations of cholesterol and low-density lipoprotein-cholesterol concentrations, and increased number of apo-B-carrying lipoproteins⁵⁰.

Prospective population studies confirm a close association between increasing body fatness and type 2 diabetes. In the Nurses Cohort Study, BMI was the dominant predictor of the risk of diabetes after adjustment for age⁵¹. The risk of diabetes increased fivefold for those with a BMI of 25, 28-fold for those with BMI of 30, and 93-fold for those women with a BMI of 35 or greater, compared with women with a BMI of less than 21. Women who gained 8–10.9 kg in weight during the period of study had a 2.7-fold increased risk of diabetes compared with women of stable weight. Similarly, the risk of diabetes

in men increases for all BMI levels of 24 or above. Compared with men with a BMI of less than 21, the risk of diabetes, adjusted for age, is increased 2.2-fold for a BMI between 25 and 26.9, 6.7-fold for a BMI between 29 and 30, and 42-fold for those with a BMI of 35 or greater⁵². The distribution of fat tissue is also associated independently with diabetes: a waist circumference of >40 inches (102 cm) increases the risk of diabetes 3.5-fold even after controlling for BMI⁵³.

Cardiovascular function in obesity

The effects of increased body fatness on cardiovascular function are predictable (Box 3). Total body oxygen consumption is increased as a result of an expanded lean tissue mass as well as the oxidative demands of metabolically active adipose tissue, and this is accompanied by an absolute increase in cardiac output. However, the values are within the normal range when they are normalized to body surface area⁵⁴. The total blood volume in obesity is increased in proportion to body weight. This increase in blood volume contributes to an increase in the left ventricular preload and an increase in resting cardiac output⁵⁵. The increased demand for cardiac output is achieved by an increase in stroke volume while the heart rate remains comparatively unchanged. The obesity-related increase in stroke volume results from an increase in diastolic filling of the left ventricle⁵⁶. The volume expansion and increase in cardiac output lead to structural changes of the heart, and the increase in left ventricular filling results in an increase in the left ventricular cavity dimension and

an increase in wall stress. As left ventricular dilatation is accompanied by myocardial hypertrophy, the ratio between ventricular cavity radius and wall thickness is preserved, and this thickening of the wall with dilatation results in eccentric hypertrophy. Left ventricular mass increases directly in proportion to BMI or the degree of overweight⁵⁷. The blood pressure is a function of cardiac output and systemic vascular resistance (the vascular resistance against which the blood is pumped). An elevated cardiac output is common with moderate obesity but not all obese patients are hypertensive. However, in those subjects where systemic resistance is increased, the combination of hypertension and obesity results in an increase of ventricular wall dimensions disproportionate to the chamber radius and this leads, in time, to concentric hypertrophy⁵⁸.

The cardiovascular adaptation to the increased intravascular volume of obesity may not completely restore normal haemodynamic function. Marked systolic dysfunction occurs when the ventricle can no longer adapt to volume overload. Dilatation of the left ventricle cavity radius leads to a decline in ventricular contractility. Despite an elevation of cardiac output, obese individuals have been shown to have depressed myocardial contractility proportional to excess weight⁵⁹. With left ventricular hypertrophy, reduced ventricular compliance alters the ability of the chamber to accommodate an increased volume during diastole and this results in diastolic dysfunction. A combination of systolic and diastolic dysfunction progresses to clinically significant heart failure. Body weight, independent of several traditional risk factors, was directly related to the development of congestive cardiac failure in the Framingham Heart Study⁶⁰.

In addition to congestive cardiac failure, the presence of left ventricular hypertrophy has been associated with a greater risk of morbidity and mortality from CHD and sudden death, as well as abnormal heart rhythms (or arrhythmias). In the Nurses Cohort Study the risk of CHD increased twofold for women with a BMI between 25 and 28.9, and 3.6-fold for a BMI >29, compared with women with a BMI of less than 21 (ref. 61). In the Framingham Heart Study, the 26-year incidence of CHD in women and men was related proportionately to excess weight. In this study the incidence of CHD increased by a factor of 2.4 in obese women and by a factor of 2.0 in obese men under the age of 50 years⁶⁰. The independent risk of CHD attributed to obesity in multivariate analysis may reflect other important mediators such as upper body fat, altered rheology and haemostasis, hyperinsulinaemia or sleep apnoea.

Sleep-breathing abnormalities in obesity

An increased amount of fat in the chest wall and abdomen has a predictable effect on the mechanical properties of the chest and the diaphragm and leads to an alteration of respiratory excursions during inspiration and expiration, reducing lung volume and altering the pattern of ventilation to each region. In addition, the increased mass of fat leads to a decrease in compliance of the respiratory system as a whole. All of these changes are significantly exaggerated when an obese person lies flat. The mass loading effect of fat requires an increased respiratory muscle force to overcome the excessive elastic recoil and an associated increase in the elastic work of breathing. The obesity-related changes in respiratory function are most important during sleep^{62,63} (Box 4).

During rapid eye movement (REM) sleep, there are decreases in voluntary muscle tone with reduced arterial oxygen saturation and a rise in carbon dioxide. These changes affect all individuals but are especially marked in obese subjects. Irregular respiration and occasional apnoeic episodes often occur in lean people during REM sleep, but obesity, with its influence on respiratory mechanics, increases their frequency and may result in severe hypoxia with resultant cardiac arrhythmias. Studies of obese men and women have demonstrated that the obstruction occurs in the larynx and is associated with loss of tone of the muscles controlling tongue movement. Relaxation of the genioglossus muscle allows the base of the tongue to

fall back against the posterior pharyngeal wall occluding the pharynx. This results in a temporary cessation of breathing (apnoea) and associated transient fall in arterial oxygen saturation concentration (hypoxia). It is not uncommon to observe low oxygen saturation values during REM sleep in some obese men while their awake arterial gases are normal⁶⁴. By contrast, premenopausal obese women show relatively minor alterations during sleep with a decrease in arterial oxygen saturation of less than 7% without apnoea. After the menopause, the changes seen in obese women become more marked with the reduction in oxygen saturation during sleep being >7% and being accompanied by apnoeic episodes⁶⁵. A minority of obese patients develop a situation characterized by a marked depression in both carbon dioxide (hypercapnic) and hypoxic respiratory drives, accompanied by abnormal and irregular pattern of breathing during sleep and (eventually) in the waking state⁶⁶. Characteristically, such individuals show frequent and prolonged episodes of sleep apnoea: sleep is disturbed with frequent awakening related to the resumption of breathing after an apnoeic episode. Daytime somnolence soon intervenes and is accompanied by persistent hypoxia/hypercapnia, pulmonary hypertension (superimposed upon an increased circulatory volume) and right-sided cardiac failure. Such changes constitute the clinical manifestation of the obesity-hypoventilation syndrome (formerly known as the Pickwickian syndrome).

In the Swedish Obesity Subjects study, which examined 3,034 subjects with a BMI above 35, over 50% of men and one-third of women reported snoring and apnoea. In contrast, 15.5% of Swedish men of comparable age were self-reported habitual snorers⁶⁷.

Several groups have reported an increased risk of myocardial infarction and stroke in sleep apnoea. Snoring is a strong risk factor for sleep-related strokes, whereas symptoms of sleep apnoea increase the risk for cerebral infarction⁶⁸.

Future prospects

It is noticeable, working in east London, how the current generation of teenage Asians is much taller and more sturdily built compared with their parents. This observation is explained largely by improved nutrition. In contrast, Asian parents are becoming obese, a situation not seen 10 years ago, and are paying a serious medical penalty as a consequence. This change in population anthropometry is not restricted to east London or to a particular ethnic group, but reflects a major global shift in body size. A sudden disproportionate rise in the number of people who are seriously obese is observed as the mean weight of a population rises⁶⁹; a situation now faced by most developed and many developing nations.

The accompanying reviews in this *Nature* Insight on Obesity confirm the identity of several genes involved in the development of obesity in animal models and describe central neural pathways concerned in the regulation of energy balance. Such genes and neural pathways are likely to be important in the genesis of human obesity but they should not detract from the importance of environmental factors — the epidemic of obesity witnessed during the past 20 years has emerged from a relatively constant genetic pool. For the future, priority must be given for achieving a better understanding of susceptible genotypes for obesity and identifying different obese phenotypes. The latter should enable particular treatments to be targeted at appropriate individuals who are at specific medical risk.

The identification of major and minor genes involved in the aetiology and pathogenesis of obesity remains critically important for the immediate future. Nevertheless, the development and implementation of effective programmes that successfully encourage increased physical activity and healthy eating across populations remain paramount for the prevention of obesity and its associated diseases — this will require the active engagement of individuals and their governments. □

1. Kopelman, P. G. Causes and consequences of obesity. *Med. Int.* 22, 385–388 (1994).

2. Society of Actuaries. Build Study of 1979 (Recording and Statistical Corporation, 1980).

3. Lew, E. A. Mortality and weight: insured lives and the American Cancer Study. *Annals of Intern. Med.* **103**, 1024–1029 (1985).
4. Hubert, H. B. The importance of obesity in the development of coronary risk factors and disease: the epidemiological evidence. *Annu. Rev. Public Health* **7**, 493–502 (1986).
5. World Health Organization Expert Committee. *Physical Status: The Use and Interpretation of Anthropometry*, WHO Tech. Rep. Ser. no. 854 (World Health Organization, Geneva, 1995).
6. World Health Organization. *Obesity: Preventing and Managing the Global Epidemic* (World Health Organization, Geneva, 1997).
7. Manson, J. E. *et al.* Body weight and mortality among women. *N. Engl. J. Med.* **333**, 677–685 (1995).
8. Willett, W. C. *et al.* Weight, weight change, and coronary heart disease in women: risk within the “normal” weight range. *J. Am. Med. Assoc.* **273**, 461–465 (1995).
9. Willett, W. C., Dietz, W. H. & Colditz, G. A. Guidelines for healthy weight. *N. Engl. J. Med.* **341**, 427–433 (1999).
10. Han, T. S., van Leer, E. M., Seidell, J. C. & Lean, M. E. J. Waist circumference action levels in the identification of cardiovascular risk factors: prevalence study in a random sample. *Br. Med. J.* **311**, 1401–1405 (1995).
11. Seidell, J. C. The impact of obesity on health status: some implications for health care costs. *Int. J. Obesity* **19**(Suppl. 6), S13–S16 (1996).
12. McKeigue, P. M., Shah, B. & Marmot, M. G. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. *Lancet* **337**, 382–386 (1991).
13. Kotani, K. *et al.* Two decades of annual medical examinations in Japanese obese children: do obese children grow into obese adults? *Int. J. Obesity* **21**, 912–921 (1997).
14. Dietz, W. H. Critical periods in childhood for the development of obesity. *Am. J. Clin. Nutr.* **59**, 829–840 (1994).
15. World Health Organization MONICA Project. Geographical variation in the major risk factors of coronary heart disease in men and women aged 35–64 years. *World Health Stat. Quart.* **41**, 115–140 (1988).
16. Prescott-Clarke, P. & Primatesta, P. *Health Survey for England 1996* (HMSO, London, 1998).
17. Kuczmarski, R. J. *et al.* Increasing prevalence of overweight among US adults. The National Health and Nutrition Examination Surveys 1960 to 1991. *J. Am. Med. Assoc.* **272**, 205–211 (1994).
18. Forrester, T. *et al.* in *The Origins and Consequences of Obesity* (eds Chadwick, D. J. & Cardeau, G.) 17–31 (Wiley, Chichester, 1996).
19. Popkin, B. M. The nutrition transition in low-income countries — an emerging crisis. *Nutr. Rev.* **52**, 285–298 (1994).
20. Hodge, A. M. *et al.* Prevalence and secular trends in obesity in Pacific and Indian Ocean island populations. *Obesity Res.* **3**(Suppl. 2), 77s–87s (1995).
21. Musaiger, A. in *Diet-Related Non-communicable Diseases in the Arab Countries of the Gulf*. (eds Musaiger, A. O. & Miladi, S. S.) 99–117 (Food and Agriculture Organisation of the United Nations, Cairo, 1996).
22. Bouchard, C. *et al.* The response to long term overfeeding in identical twins. *N. Engl. J. Med.* **322**, 1477–1482 (1990).
23. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
24. Campfield, L. A., Smith, F. J., Guisey, Y., Devos, R. & Burn, P. Recombinant mouse ob protein: evidence for a peripheral signal linking adiposity and central neural networks. *Science* **280**, 546–549 (1995).
25. Maffei, M. *et al.* Leptin levels in human and rodent: measurement of plasma leptin and ob RNA in obese and weight reduced subjects. *Nature Med.* **11**, 1155–1161 (1995).
26. Farooqi, I. S. *et al.* Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *N. Engl. J. Med.* **341**, 879–884 (1999).
27. Prentice, A. M., Black, A. E., Coward, W. A. & Cole, T. J. Energy expenditure in affluent societies: an analysis of 319 doubly-labelled water measurements. *Eur. J. Clin. Nutr.* **50**, 93–97 (1996).
28. Ravussin, E. *et al.* Reduced rate of expenditure as a risk factor for body weight. *N. Engl. J. Med.* **318**, 467–472 (1988).
29. Seidell, J., Muller, D., Sorkin, J. & Andres, R. Fasting respiratory exchange ratio and resting metabolic rates as predictors of weight gain: The Baltimore Longitudinal Study on Ageing. *Int. J. Obesity* **16**, 667–674 (1992).
30. Ferro-Luzzi, A. & Martino, L. in *The Origins and Consequences of Obesity* (eds Chadwick, D. & Cardew, G.) 207–227 (Wiley, Chichester, 1996).
31. Rissanen, A. *et al.* Determinants of weight gain and overweight in adult Finns. *Eur. J. Clin. Nutr.* **45**, 419–430 (1991).
32. Gormaker, S. *et al.* Television viewing as a cause of increasing obesity among children in the United States. *Arch. Pediatr. Adolesc. Med.* **150**, 356–362 (1996).
33. Prentice, A. M. & Jebb, S. A. Obesity in Britain: gluttony or sloth? *Br. Med. J.* **311**, 437–439 (1995).
34. Lichtman, S. *et al.* Discrepancy between self-reported and actual calorie intake and exercise in obese subjects. *N. Engl. J. Med.* **327**, 1893–1898 (1993).
35. Poppitt, S. D., Swann, D., Black, A. E. & Prentice, A. M. Assessment of selective under-reporting of food intake by both obese and non-obese women in metabolic facility. *Int. J. Obesity* **22**, 303–311 (1998).
36. Lawton, C., Burley, V., Wales, J. & Blundell, J. Dietary fat and appetite control in obese subjects: weak effects on satiation and satiety. *Int. J. Obesity* **17**, 409–416 (1993).
37. Blundell, J. E. & Macdiarmid, J. I. Passive overconsumption: fat intake and short term energy balance. *Ann. NY Acad. Sci.* **827**, 392–407 (1997).
38. James, W. in *The Origins and Consequences of Obesity* (eds Chadwick, D. & Cardew, G.) 1–16 (Wiley, Chichester, 1996).
39. Ravussin, E. Metabolic differences and the development of obesity. *Metabolism* **9**(Suppl. 3), 12–14 (1995).
40. Wilks, R. *et al.* in *The Origins and Consequences of Obesity* (eds Chadwick, D. & Cardew, G.) 37–53 (Wiley, Chichester, 1996).
41. Bennett, N., Dodd, T., Flatley, J., Freeth, S. & Bolling, K. *Health Survey for England 1993* (HMSO, London, 1995).
42. Barker, D. J. P. Fetal origins of coronary heart disease. *Br. Med. J.* **311**, 171–174 (1995).
43. Barker, D. J. P. *et al.* Growth in utero, blood pressure in childhood and adult life, and mortality from cardiovascular disease. *Br. Med. J.* **298**, 654–657 (1989).
44. Hales, C. N. *et al.* Fetal and infant growth and impaired glucose tolerance at age 64. *Br. Med. J.* **303**, 1019–1022 (1991).
45. Ravelli, J. P., Stein, Z. A. & Susser, M. W. Obesity in young men after famine exposure in utero and early infancy. *N. Engl. J. Med.* **295**, 349–353 (1976).
46. Kolterman, O. G., Insel, J., Sackow, M. & Olefsky, M. Mechanisms of insulin resistance in human obesity. *J. Clin. Invest.* **65**, 1272–1284 (1980).
47. Reynisdottir, S., Ellerfeldt, K., Wahrenberg, H., Lithell, H. & Arner, P. Multiple lipolysis defects in insulin resistance [metabolic] syndrome. *J. Clin. Invest.* **93**, 2590–2599 (1994).
48. Lonnquist, F., Thorne, A., Nilsell, K., Hoffstedt, J. & Arner, P. A pathogenic role of visceral fat β_2 -adrenoceptors in obesity. *J. Clin. Invest.* **95**, 1109–1116 (1995).
49. Frayn, K. N., Williams, C. M. & Arner, P. Are increased plasma non-esterified fatty acid concentrations a risk marker for coronary heart disease and other chronic diseases? *Clin. Sci.* **90**, 243–253 (1996).
50. Sniderman, A. & Cainflone, K. Metabolic disruptions in the adipocyte-hepatocyte fatty acid axis as the cause of hyperapoB. *Int. J. Obesity* **19**(Suppl. 1), S27–S33 (1995).
51. Colditz, G. A., Willett, W. C., Rotnitsky, A. & Manson, J. E. Weight gain as a risk factor for clinical diabetes in women. *Arch. Int. Med.* **122**, 481–486 (1995).
52. Chan, J. M. *et al.* Obesity, fat distribution and weight gain as risk factors for clinical diabetes in men. *Diabetes Care* **17**, 961–969 (1994).
53. Lean, M. E. J., Hans, T. S. & Seidell, J. C. Impairment of health and quality life in people with large waist circumference. *Lancet* **351**, 853–856 (1998).
54. Masserli, F. H. *et al.* Borderline hypertension and obesity: two prehypertensive states with elevated cardiac output. *Circulation* **66**, 55–60 (1982).
55. De Divitiis, O. *et al.* Obesity and cardiac function. *Circulation* **64**, 477–482 (1981).
56. Licata, G. *et al.* Effect of obesity on left ventricular function studied by radionuclide angiography. *Int. J. Obesity* **15**, 295–302 (1991).
57. Lauer, M. S., Anderson, K. M., Kannel, W. B. & Levy, D. The impact of obesity on left ventricular mass and geometry. *J. Am. Med. Assoc.* **266**, 231–236 (1991).
58. de la Maza, M. *et al.* Ventricular mass in hypertensive and normotensive obese subjects. *Int. J. Obesity* **18**, 193–197 (1994).
59. Garavaglia, G. E. *et al.* Myocardial contractility and left ventricular function in obese patients with essential hypertension. *Am. J. Cardiol.* **62**, 594–597 (1988).
60. Hubert, H. B., Feinleib, M., McNamara, P. M. & Castelli, W. P. Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham heart study. *Circulation* **67**, 968–977 (1983).
61. Willett, W. C. *et al.* Weight, weight change and coronary heart disease in women. *J. Am. Med. Assoc.* **273**, 461–465 (1995).
62. Kopelman, P. G. Sleep apnoea and hypoventilation in obesity. *Int. J. Obesity* **16**(Suppl. 2), S37–S42 (1992).
63. Grunstein, R. R. in *Clinical Obesity* (eds Kopelman, P. G. & Stock, M. J.) 248–289 (Blackwell Science, Oxford, 1998).
64. Kopelman, P. G., Apps, M. C. P., Cope, T. & Empey, D. W. Nocturnal hypoxia and sleep apnoea in asymptomatic men. *Int. J. Obesity* **10**, 211–217 (1986).
65. Kopelman, P. G., Apps, M. C. P., Cope, T. & Empey, D. W. The influence of menstrual status, body weight and hypothalamic function on nocturnal respiration in women. *J. R. Coll. Physicians Lond.* **19**, 243–247 (1985).
66. Rochester, D. F. & Eaton, Y. Current concepts in the pathogenesis of the obesity-hypoventilation syndrome. Mechanical and circulatory factors. *Am. J. Med.* **57**, 402–420 (1974).
67. Grunstein, R. R., Stenlof, K., Hedner, J. & Sjostrom, L. Impact of obstructive apnoea and sleepiness on metabolic and cardiovascular risk factors in the Swedish Obese Subjects (SOS) Study. *Int. J. Obesity* **19**, 410–418 (1995).
68. Palomaki, H., Partinen, M., Erkinjuntti, T. & Kaste, M. Snoring, sleep apnoea syndrome and stroke. *Neurology* **42**(Suppl. 6), 75–81 (1992).
69. Rose, G. Population distribution of risk and disease. *Nutr. Metabol. Cardiovascular Dis.* **1**, 37–40 (1991).

Genetics of body-weight regulation

Gregory S. Barsh*, I. Sadaf Farooqi† & Stephen O'Rahilly†

*Departments of Pediatrics and Genetics and the Howard Hughes Medical Institute, Beckman Center, Stanford, California 94305-5428, USA

†Departments of Medicine and Clinical Biochemistry, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK

The role of genetics in obesity is twofold. Studying rare mutations in humans and model organisms provides fundamental insight into a complex physiological process, and complements population-based studies that seek to reveal primary causes. Remarkable progress has been made on both fronts, and the pace of advance is likely to accelerate as functional genomics and the human genome project expand and mature. Approaches based on mendelian and quantitative genetics may well converge, and lead ultimately to more rational and selective therapies.

The application of genetics to understand and treat disorders of human body-weight regulation has made remarkable progress in recent years. Most of the previously existing mutations in mouse obesity genes that segregate as mendelian traits have now been cloned, and several homologous mutations have been discovered as rare causes of human obesity. Phenotypes shared between mice and humans with homologous obesity mutations show deep conservation of the underlying pathways; at the same time, particular features in some of the human patients reveal aspects of energy homeostasis that are unique to human physiology. In nearly every case, these genes and their products have provided molecular focal points for new biochemical and physiological pathways that are conserved among all mammals. Basic insight is sorely needed because, unlike other common human diseases such as diabetes or hypertension, there have until recently been few targets for intervention.

To investigate common variation in body weight and adiposity, new methods of quantitative genetics have been applied to human populations, and to inbred strains of mice that vary in their body composition. These studies have led to the first genetic maps for obesity in humans and mice, and have helped to confirm that susceptibility to obesity is controlled largely by genetic factors.

Nonetheless, the quantitative genetic and mendelian

approaches so far show little overlap and, in fact, may never fully converge. With one or two notable exceptions, the map positions of obesity loci identified by quantitative studies do not correspond to 'classical' obesity mutations such as *ob*, *tubby* or *fat*. This apparent lack of concordance may reflect a relatively large pool of genes in which allelic variation has the potential to impact on body-weight regulation. Our review will summarize results from both quantitative genetic and mendelian approaches, emphasizing the differences in primary goals and the potential ways in which each type of approach might improve human health.

Interplay between genetics and environment

Body size in laboratory animals is the archetypal polygenic trait, a quantitative phenotype that fails to display a mendelian inheritance pattern because it is controlled by many different loci¹. Identifying such loci for human obesity is further confounded by the effects of environmental variables. Nonetheless (and perhaps surprisingly), pioneering studies²⁻⁶ have indicated that genetic factors account for a substantial portion of variation in human adiposity. In a review⁷ of studies that totalled more than 100,000 individuals from biological and adoptive relationships, correlations of body-mass index (BMI; defined as weight in kilograms divided by the square of the height in metres) were determined for monozygotic and dizygotic twin pairs,

Table 1 Loci identified by genome-wide linkage studies as contributing to human obesity or related intermediate phenotypes

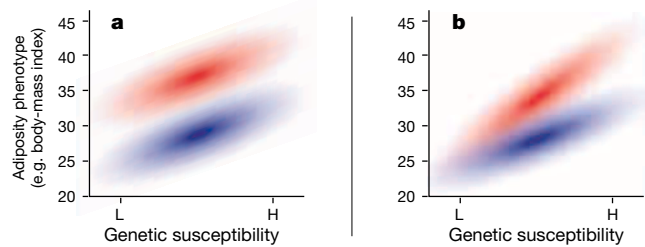
Chromosome location	Phenotype	Design	Lod score	Population	Candidate genes	Reference
2p	Plasma leptin	Extended pedigree	4.95	Mexican-American	POMC	15
2p	Plasma leptin	Sib pair	2.68	French	POMC	16
10p	BMI	Sib pair	4.85	French	None	16
5p	Plasma leptin	Sib pair	≈3.0	French	CART	16
11q	Percentage body fat	Sib pair	2.8	Pima Indians	None	104
11q	BMI	Sib pair	3.5	Pima Indians	None	105
20q	BMI, percentage body fat	Extended pedigree and sib pair	3.06–3.17	US Caucasians	ASIP, CEBPB, GNAS1	106

Loci with a reported lod score of 2.6 or greater are included. Major candidate genes known to be located within the region of linkage are indicated. The two studies describing 11q linkage to percentage body fat and BMI are not fully independent.

Box 1

Interaction between genes and environment in obesity

The figure opposite depicts genetic susceptibility as a conceptualized variable that measures allelic variation at a common set of polygenic obesity genes in low- and high-risk environments, indicated by blue and red, respectively. For example, in an environment where caloric availability is limited (blue), individuals with high genetic susceptibility might have a commensurately higher degree of adiposity, but the mean level would still be relatively low. The same pool of alleles in an environment replete with high-fat or high-calorie diets (red) would shift the overall distribution of adiposity upward such that high genetic susceptibility would lead to morbid obesity (panel **a**). Perhaps a more likely scenario, however, are gene $\times$ environment interactions whereby a high level of genetic susceptibility would have little effect on adiposity except in a predisposing environment (panel **b**). An effect of caloric availability on phenotypic expression is significant in a historical context or with regard to developing-country populations, whereas differences in the prevalence of obesity among Western society today is more likely explained by cultural and socioeconomic variables such as meal pattern, television exposure and physical activity.



Box 1 Figure Potential effects of genes and environment on adiposity. The intensity of each colour corresponds to the population frequency. **a**, In the absence of gene $\times$ environment interactions, environmental risk factors have a large effect on phenotype, but the effect is independent of genetic factors. **b**, In the presence of gene $\times$ environment interactions, the effects of high genetic susceptibility are amplified by a high-risk environment.

and for biological parent-offspring and adoptive relative pairs. A weighted mean BMI correlation of 0.74 was obtained for monozygotic twins compared with a BMI correlation of 0.32 for dizygotic twin pairs, which corresponds to an estimated heritability of between 50 and 90%. Furthermore, correlations for biological parent-offspring pairs and adoptive relative pairs were 0.19 and 0.06, respectively, which indicates a relatively minor role for cultural transmission.

However, a genetic aetiology for common human obesity is difficult to reconcile with the marked variation in prevalence observed as a function of socioeconomic and/or demographic factors^{8,9}. Obesity is largely a disease of post-industrial society, in which substantial variations in prevalence still exist that point to cultural factors. In Sweden, for example, the prevalence of obesity is less than half that of the United States, where adiposity is increasing at an alarming rate, especially in children⁹.

To some extent, conclusions based on genetic and epidemiological studies are limited to the populations in which they have been carried out, and it is possible that the relative role for environmental factors will loom increasingly large in those populations where obesity is especially prevalent. A more parsimonious view, however, is based on studies of laboratory rodents¹⁰ (see below), in which the effect of genetic background on the response to high-calorie or high-fat diets is equally, if not more striking, than the effect of genetic background on adiposity when food intake is held constant. Thus, one set of genes controls efficiency of caloric utilization, whereas another set — not necessarily mutually exclusive — are recognized by their effects on diet-induced obesity. It would be surprising if a parallel situation did not exist in humans.

These considerations support a model in which susceptibility to obesity is determined largely by genetic factors, but the environment determines phenotypic expression (Box 1). The concept that 'nurture' operates on an underlying pool of genes that contribute to obesity susceptibility has important implications for our approach to the prevention and treatment of obesity. If some environmental variables manifest themselves only on certain genotypes, efforts to prevent obesity at a public health level can be focused on recognition and counselling of susceptible individuals. In addition, appreciating the importance of genetic variation as an underlying cause helps to dispel the notion that obesity represents an individual defect in behaviour with no biological

basis, and provides a starting point for efforts to identify the genes involved.

The search for human obesity genes

Genetic determinants of inter-individual variation in obesity and related phenotypes are likely to be multiple and interacting, with most single variants producing only a moderate effect. To recognize such effects, a common approach is to measure multiple intermediate components where the effects of individual genes may be more apparent — for example, levels of leptin (the adipocyte-derived hormone that has a central role in regulating energy balance), percentage body fat, or fat distribution — and test each of these phenotypes independently for co-segregation with polymorphic markers spaced at regular intervals throughout the genome¹¹. This approach casts a wider net, for example, for genes that control leptin independently of body fat. It also has important practical implications as the ratio of visceral to subcutaneous adiposity affects morbidity independent of overall BMI¹². Theoretically, genetic interactions can be recognized by testing for simultaneous association between multiple loci and the disease phenotype, but in practice, this approach is more easily applied to experiments based on inbred strains of mice¹³ (see below). In either case, however, one quickly runs into the problem of multiple testing, that is, the likelihood that a specific association will exceed a nominal or pointwise significance level increases according to the number being tested¹⁴.

For example, the traditional lod-score threshold of 3 for linkage between one phenotype and one locus corresponds to a nominal *P* value of 0.0001, but when multiple loci spanning the genome are tested simultaneously, the overall or 'genome-wide' significance level is only 9%. Spurred, in part, by the need "to avoid a flood of false-positive claims", Lander and Kruglyak¹⁴ recommended adopting lod-score thresholds of 3.3–3.8 (depending on study design) for declaring "significant linkage", that is, a genome-wide significance level $\leq 5\%$. These calculations take into account simultaneous testing of multiple loci, but not multiple phenotypes; the latter is a more difficult problem as, in most cases, phenotypes such as leptin and adiposity will be correlated and not independent.

Thorny statistical issues notwithstanding, a common human obesity gene identified by linkage is especially convincing when replicated in an independent study. Results from four reported genome-wide linkage studies that examined obesity and/or related intermediate traits have identified several loci that show positive

evidence for linkage (Table 1). In two studies, one of extended, Mexican-American pedigrees¹⁵ and the other of French sibling pairs¹⁶, significant linkage of serum leptin levels to the short arm (p) of chromosome 2 at band 21 (2p21) was found. In the Mexican-American study, evidence indicating linkage of fat mass to 2p21 was reported, whereas in the French study where a BMI above 27 was used to define a qualitative trait, there was no evidence of linkage to 2p21. However, the potential importance of this locus is supported by a recent study in which linkage to serum leptin levels was confirmed in a population of African-Americans¹⁷. Notably, human chromosome 2p21 includes the pro-opiomelanocortin (POMC) gene in which a complete loss-of-function causes monogenic obesity in mice and humans (see below). By using single-nucleotide polymorphisms (SNPs) around the POMC gene, an association between extended haplotypes and serum leptin levels has been found in Mexican-Americans¹⁸; but the degree of the association is small, which indicates that multiple alleles or possibly other loci are also involved.

Evidence supporting some of the loci described above is strong, but other than 2p21, none has yet been replicated in independent populations. One possible interpretation is that the genetic determinants of BMI and related traits are markedly different between ethnic groups. Alternatively, if the polygenic predisposition to obesity is determined by many genes that each have a small effect, it may be difficult to reach the thresholds mentioned above in a single study. To avoid "the risk of causing true hints of linkage to be missed", Lander and Kruglyak¹⁴ also proposed thresholds for declaring "suggestive linkage" that correspond to an expected incidence, for a single phenotype, of one false positive per genome scan. Communication of suggestive linkage among investigators in the field can help guide the design of future studies while still maintaining rigour and credibility. One such forum, the 'human obesity gene map', has been established¹⁹ and contains entries for more than 40 genes and 15 chromosomal regions in which published studies indicate a possible relationship to adiposity or a related phenotype. Entries in the human obesity gene map are based not only on genome-wide linkage scans, but also on candidate genes tested for linkage and/or association (see below). What may ultimately prove most useful, however, are mechanisms to pool data for meta-analyses¹¹. Combining the results of genome scans will be especially powerful in the next few years as SNPs and efficient parallel-detection methods are developed²⁰.

Polygenic obesity in rodents as a model system

Many of the difficulties inherent in genetic analysis of human obesity can be minimized by using a model organism in which the effects of

environmental and genetic background are held constant. Inbred strains of mice are especially useful because different strains encompass a wide range of adiposity and growth, and because there exists an excellent set of genomic tools for molecular analysis of quantitative trait loci (QTLs)^{21–23}. In most cases, genetic differences between 'obese' and 'non-obese' strains represent random assortment of alleles that occurred during the course of inbreeding, but a few strains and outbred stocks have been developed from selection experiments that were directed specifically at altering body size or energy balance^{24,25}. For example, the SM/J and LG/J strains, which were developed independently more than 50 years ago^{26,27}, exhibit mean interstrain body-weight differences nearly 10 times that of the intrastrain standard deviation. Analysis of SM/J and LG/J intercross mice should reveal not only the number, size of effect, and epistatic relationships of QTLs, but also how interaction between different loci is affected by population bottlenecks or selection^{28,29}.

A particular advantage of using mice or other laboratory animals as a model system in quantitative genetics is the opportunity to target specific phenotypes such as energy expenditure or gene × environment interactions. Mice selected for high heat loss are leaner, but, perhaps surprisingly, hyperphagic, compared with those selected for low heat loss²⁵. QTL analysis offers the opportunity to determine whether some genes control energy expenditure independently of hyperphagia and adiposity, while others act coordinately on all components of the energy balance equation. From the perspective of human obesity, perhaps the most important environmental variable is diet composition. Studies indicate that small differences in adiposity among inbred strains are magnified by a high-fat diet^{10,30–32}, providing support for the model depicted in panel b of the figure in Box 1.

Ironically, the relative ease with which obesity QTLs can be identified in mice compared to humans has led to a mouse gene map for polygenic obesity that is highly reliable yet extremely dense¹⁹. Over 70 loci have been identified from more than 15 different crosses; many of these pass statistical muster and some have been replicated independently in the context of the same strain pair (Table 2). In general, the patterns of loci identified in different strain pairs show little overlap, which indicates that the field is far from saturation and, consequently, there are likely to be many genes in which allelic variation can account for body-weight differences between two populations. Because confidence limits for the mouse obesity QTLs are fairly broad, homologous locations in the human genome are extensive (Fig. 1), and the question of whether the actual genes in mice are homologous to those in humans remains unanswered. Molecular identification may help, but this is not so straightforward

Table 2 Genome-wide linkage studies for polygenic obesity in different mouse strains

Strain pair	Phenotype	No. loci	Chromosome location	Reference
QS × C57BL/6J	Body weight	1	10	107
AKR/J × SWR/J	Diet-induced obesity	3	4, 9, 15	30,32
C57BL/6J × SPRET	Adiposity, fat distribution	4	6, 7, 12, 15	108
(A/J × SPRET) × C57BL/6J	Body weight	3	X	109
129/Sv × Le/Suz	Adiposity	2	1, 7	110
LG/J × SM/J	Body weight	12	1, 2, 3, 4, 6, 7, 9, 11, 12, 13, 14, 18	29
C57BL/6J × CAST/Ei	Diet-induced obesity	1	15	111
C57BL/6J × DBA/2J	Body weight	11	1, 4, 5, 6, 7, 9, 11, 13, 14, 17	112
AKR/J × C57L/J	Adiposity	2	2, 17	113
NZB/B1NJ × SM/J	Adiposity	1	2	114
C57BL/6J × CAST/Ei	Adiposity, body weight, fat distribution	5	2, 4, 9, 15	115
DU6 × DUK	Body weight	5	3, 4, 11, 13	24
KK × C57BL/6J — A/a	Adiposity, body weight	2	4, 6	116
KK × C57BL/6J	Adiposity	2	X, 9	117
MH × C57BL/6J	Adiposity, body weight, heat loss	17	1, 2, 3, 7, 11, 17	25

The DU6, DUK, QS and MH animals are heterogeneous populations selected for obesity-related traits. Most of the entries listed above evaluate adiposity and fat distribution as percentage of body weight accounted for by various fat depots; some studies evaluate total body weight and therefore survey general aspects of growth and nutrition. In general, included loci show peak lod scores >3.3; homologous locations in the human genome are shown in Fig. 2 and in ref. 19, from which some of these data are abstracted.

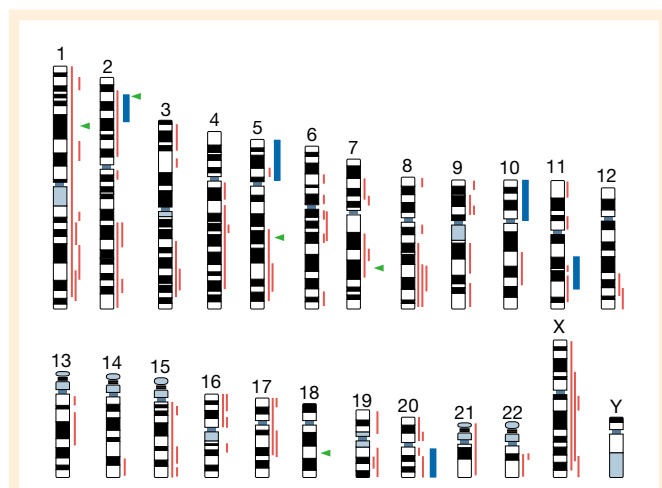


Figure 1 Chromosomal location of obesity genes. Ideogram of human karyotype with human monogenic mutations (Table 3) indicated in green, human loci identified by genome-wide linkage scans (Table 1) indicated in blue, and the potential location of one or more mouse QTLs (based on conserved homology) indicated in red, as described in Table 2 of ref. 19. The potential mouse QTL locations correspond to a large fraction of the human genome because there are many, and confidence limits for QTL mapping usually contain multiple segments of conserved homology.

for QTLs, the effect of which may depend on other genes, or for those identified in highly divergent strain pairs. Nonetheless, a quantitative genetic approach to body weight in mice offers important insight into fundamental aspects of genetic architecture that can be addressed only in a model system.

Biological insight from obesity mutations in mice

Much of the recent excitement about understanding and treating disorders of human body-weight regulation is based on the identification of genes responsible for previously existing mouse obesity mutations, and the subsequent realization that several of these genes uncover fundamental physiological pathways that were unappreciated before then (Fig. 2). Over the past hundred years, the *A^y*, *db*, *fat*, *ob* and *tub* mutations were each recognized as spontaneous variants on the background, or during the derivation, of modern inbred strains of mice. Because the mutant animals have been substrates for a long series of endocrinological, metabolic and behavioural investigations, the fruits of positional cloning realized over the past few years have provided considerable insight into the biology that underlies each mutation^{33,34}.

This has been most pronounced in the case of *ob* and *db*, for which pioneering studies³⁵ had predicted an adiposity-sensing pathway regulated by a circulating hormone and its cognate receptor, now known as leptin and Ob-R (leptin receptor), respectively (see review by Schwartz *et al.*, pp. 661–671). Production and/or release of leptin by adipocytes serves as the afferent component of a regulatory loop that regulates energy balance through behavioural and metabolic effectors. Recognition of this pathway was an immediate product of cloning the gene³⁶, and has been a watershed event in mammalian physiology, seeding activity in arenas of basic science, clinical investigation and the pharmaceutical industry³⁷.

The central actions of leptin can be viewed as one arm of a homeostatic circuit or, alternatively, as a tonic signal whose absence triggers a series of neuroendocrine responses that conserve energy when food availability is limited^{38–40}. In either case, considerable attention is now focused on deciphering the neural pathways that coordinate behavioural and metabolic effects that lie downstream of leptin. An important clue has come from the *A^y* mutation, in which an unusual genomic rearrangement leads to ectopic expression of the agouti coat-colour protein, causing a dominantly inherited

syndrome of obesity, increased growth and yellow hair colour^{41,42}. Pharmacological and transgenic experiments have revealed that obesity in *A^y* mutant mice reflects the ability of agouti protein abnormally expressed in the brain to mimic the neuropeptide agouti-related protein (Agrp), which is normally expressed in the hypothalamus and signals through the central nervous system (CNS)-specific melanocortin-4 receptor (Mc4r)^{43–45}. Agouti protein and Agrp are antagonistic ligands for melanocortin receptors. However, these receptors were recognized by, and named after, products derived from the *Pomc* gene such as α -melanocyte stimulating hormone (α -MSH) and adrenocorticotrophic hormone (ACTH), which function as agonist ligands of melanocortin receptors. The genetic predictions of these relationships have been borne out by transgenic experiments: gain-of-function Agrp mutations produce an obesity phenotype similar to that displayed by loss-of-function mutations in *Pomc* or the *Mc4r* (refs 46–48).

As described elsewhere and illustrated in Fig. 2, *Pomc*- and *Agrp*-expressing neurons in the arcuate nucleus of the hypothalamus may act as initial processors in the CNS that integrate peripheral information conveyed, in part, by circulating leptin^{44,49–55}. Although leptin clearly has some effects that are independent of melanocortin signalling^{56,57}, it is quite striking that three of five previously existing obesity mutations in mice, *A^y*, *db* and *ob*, have uncovered different nodes of the same general pathway.

As genotype–phenotype correlations made through gene targeting have caught up and surpassed those made through positional cloning, the importance of the melanocortin pathway has been reinforced by the results of *Mc4r* and *Pomc* knockouts, which produce a pattern of obesity and metabolic derangement similar to that displayed by *A^y* mice^{48,58}. Gene targeting has also identified entry points for several new pathways that are important in body-weight regulation (Table 3), some of which were anticipated, for example, *Htr2c* (ref. 59), and some of which were not, for example, *Vgf* (ref. 60), a neuroendocrine protein of unknown function. For the most part, obesity in mutant mice has been observed as a result of loss-of-function mutations, simply because hypomorphic or amorphic mutations generally occur more frequently (and are easier to make through gene targeting) than hypermorphic or neomorph mutations. This may help to explain why, for certain proteins now known to have a primary and key role in body-weight regulation, genotype–phenotype correlations have arisen by gene targeting rather than by spontaneous mutation. For example, melanin-concentrating hormone (Mch)-knockout mice are hypophagic and lean⁶¹, a phenotype less likely than obesity to be recognized as a mutation in a large population of inbred mice.

Rare obesity mutations in human populations

As with other complex human diseases, positional cloning of rare obesity mutations does not directly address genetic causes in the population as a whole, but, as illustrated above for mice, the technique could provide considerable insight into the underlying physiological pathways. In principle, the same approaches used to isolate *A^y*, *ob*, *fat* and *tub* could also be applied to humans. However, positional cloning of rare human obesity genes is more difficult than in mice, as a quantitative trait influenced by multiple genes and environmental factors can easily obscure extreme phenotypes that might otherwise exhibit single-gene inheritance patterns. Thus it is not surprising that most of the human obesity mutations recognized by a mendelian pattern of inheritance are pleiotropic syndromes in which obesity is one of several features. One or more map positions have been determined for the Prader–Willi⁶², Cohen⁶³, Alstrom⁶⁴, Bardet–Biedl^{65,66} and Borjeson–Forssman–Lehmann syndromes⁶⁷, but the causative genes have not yet been isolated⁶⁸.

An alternative strategy used with considerable success is to screen obese human patients for mutations in candidate genes selected on the basis of the mouse genetic studies. Obviously, this approach will succeed only for genes whose role in obesity

was previously suspected; nonetheless, finding that homologous mutations cause homologous phenotypes in mice and in humans underscores the fundamental nature of the underlying pathway, and validates that pathway as a target for potential intervention. Furthermore, differences in the extended phenotypes of humans and mice can reveal important species-specific differences in endocrine and energy physiology. Finally, the opportunity to treat certain patients with rare forms of genetic obesity has important implications for whether, and how, similar therapies might be used more widely.

Two kindreds with defects in leptin have been reported^{69,70}; patients from each are homozygous for different loss-of-function mutations and exhibit similar phenotypes of morbid obesity, increased appetite and hyperphagia, and hypogonadotropic hypogonadism. A leptin-receptor mutation has also been reported in one family⁷¹; affected individuals are homozygous for a mutation that truncates the receptor before the transmembrane domain. In the human patients, a striking difference between those with mutations in leptin compared with those having mutations in its receptor is the presence of significant growth retardation and central hypothyroidism in the latter. By contrast, mice deficient for leptin or its receptor display remarkably similar endocrine abnormalities; both exhibit reduced levels of growth hormone and retarded linear growth⁷². One plausible explanation for these differences would be if the human but not the mouse receptor were capable of stimulating a low level of certain hypothalamic-releasing hormones in the absence of leptin, such that loss-of-function in the receptor leads to a more severe defect than loss-of-function in the ligand. Comparison of leptin-signalling defects in mice and humans also reveals differences in the hypothalamus–pituitary–adrenal (HPA) axis; *db/db* and *ob/ob* mice exhibit markedly increased glucocorticoid production not seen in any of the human patients.

In addition to leptin, the importance of the melanocortin system in the control of human body weight has been shown by identification of obese patients with mutations in POMC or MC4R. Two children who are homozygous or compound heterozygous for different loss-of-function POMC mutations⁷³ exhibit a range of conditions that reflect impaired signalling through several melanocortin receptors. The children are not only obese but they also display altered pigmentation and adrenal insufficiency, which is due to absence of α -MSH and ACTH, melanocortin ligands for MC1R and MC2R, respectively. Although melanocortins and β -endorphin, which is also derived from POMC, have been implicated in many additional aspects of human biology, the two reported patients have no obvious impairment of behaviour,

cognition or immune function.

Several groups have reported MC4R mutations in populations or families that seem to cause obesity in a dominant fashion^{74–76}. The endocrinological and metabolic phenotypes are, in general, similar to those caused by impaired melanocortin signalling in mice, with moderate to severe obesity, little or no disturbance of the HPA axis, and normal neuroendocrine function with regard to growth, reproduction and thyroid function. Where tested, MC4R mutations cause a loss-of-function suggesting haploinsufficiency rather than a dominant negative mechanism^{77,78}. Haploinsufficiency is also observed in mice, where 0, 1 or 2 doses of an *Mc4r*-knockout allele cause increasingly progressive weight gain⁴⁸. By contrast to mutations in leptin, its receptor or POMC, mutations in MC4R are remarkably prevalent, found in approximately 3–5% of patients with a BMI above 40. However, not all patients with MC4R mutations are obese; in extended surveys of several families in which MC4R mutations are segregating, the range of BMI overlaps considerably for carriers compared with non-carriers, especially for males⁷⁹. In this regard, mutations in MC4R are important not only because of their prevalence, but also because the lessons learned may be useful to guide the search for additional common obesity genes.

For rare cases of patients with leptin or POMC deficiency, hormone replacement is an obvious form of treatment. One of the original probands with leptin deficiency has now been treated for more than a year with daily subcutaneous injections of recombinant leptin, and has experienced a marked reduction in body weight, adiposity, appetite and age-appropriate progression of gonadotropic function⁸⁰. This approach need not be limited to patients with monogenic obesity; for example, synthetic *Mc4r* agonists may prove useful in treating forms of obesity accompanied by decreased melanocortinergic activity regardless of the proximate cause.

To some extent, the remarkable parallels between humans and mice for monogenic obesity reflect a bias of discovery as searches for most of the human mutations have been guided by previously existing mouse mutations. On the other hand, nearly all of these searches have been successful, and, in the case of *Pomc*, description of the human mutation preceded that in the mouse. A telling example is also evident from the *fat* mutation, in which mutations in carboxypeptidase E (*Cpe*) lead to defective processing for several neuropeptides and prohormones⁸¹. Although no humans with *CPE* defects have been described so far, a related abnormality caused by a deficiency for prohormone convertase 1 (*PC1*) has been described in a single Caucasian female with severe early-onset obesity, hypogonadotropic hypogonadism, severe hyperproinsulinaemia and impaired adrenal

Table 3 Mendelian causes of obesity in humans and mice

Mutation and gene product	Genetic and biological mechanism	Method of identification	
		Human	Mouse
<i>diabetes</i> (leptin receptor)	Loss-of-function; brain and peripheral actions	Candidate (mouse phenotype) ⁷¹	Positional candidate ^{118–120}
<i>fat</i> (carboxypeptidase E)	Loss-of-function; processing and sorting receptor for endocrine and neuroendocrine prohormones	No	Positional candidate ⁸¹
5-Hydroxytryptamine receptor 2C (<i>Htr2c</i>)	Loss-of-function; brain serotonin receptor	No	Knockout ⁴⁵⁹
<i>lethal yellow</i> , <i>A'</i> (agouti protein)	Ectopic expression; secreted antagonist of melanocortin receptors	No	Positional cloning ^{121,122}
Melanocortin-4 receptor (<i>Mc4r</i>)	Loss-of-function; brain receptor for α -MSH and agouti-related protein	Candidate (mouse phenotype) ^{74–76}	Knockout ⁴⁸
Nescient basic helix–loop–helix 2 (<i>Nhlh2</i>)	Loss-of-function; transcription factor required for hypothalamic neuronal development	No	Knockout ¹²³
<i>obese</i> (leptin)	Loss-of-function; circulating hormone signalling adipose depot	Candidate ^{69,70}	Positional cloning ⁴⁶
Prohormone convertase 1 (<i>PCSK1</i>)	Loss-of-function; subtilisin-like protease for prohormones	Candidate (physiology) ⁸²	No
Pro-opiomelanocortin (<i>Pomc</i>)	Loss-of-function; precursor of brain α -MSH and β -endorphin	Candidate (physiology) ⁷³	Knockout ⁴⁵⁸
<i>tubby</i> (tubby protein)	Loss-of-function; probable transcription factor in hypothalamus and retina ¹²⁴	No	Positional cloning ¹²⁵

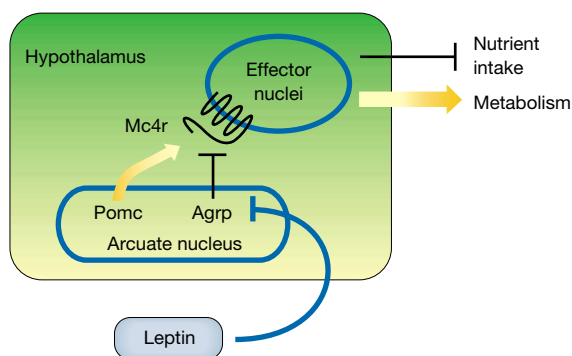


Figure 2 Action of leptin on melanocortinergic pathways. Within the hypothalamus, effector nuclei include the paraventricular nucleus and the lateral hypothalamic area. The diagram presents a simplified overview and does not include several other neuropeptides and pathways within the hypothalamus that have been implicated in body-weight regulation or leptin action. See review by Schwartz *et al.*, pp. 661–671, for additional details.

function⁸². Thus, even though *Cpe* and *PC1* code for non-homologous genes, their loss-of-function causes similar metabolic abnormalities, which in both cases cause obesity.

Genetic candidates and quandaries

The bulk of the published literature on genetic determinants of human obesity actually represents an approach that is distinct from those of linkage or mutation detection in rare mendelian disorders that are discussed above. Candidate genes are those whose dysfunction might reasonably be expected to result in obesity by virtue of their having putative or established effects on energy intake, energy expenditure or nutrient partitioning. Such effects may become apparent through our knowledge of the functional biology of their product or through the fact that major disruption of their function produces monogenic forms of obesity in humans or in animal models. Testing a particular candidate gene often consists of determining whether polymorphic markers in or around the gene are present more frequently in a group of obese compared with non-obese patients.

But this approach can be fraught with difficulties. In heterogeneous populations, one can never be certain that cases and controls come from the same genetic background. The use of family-based controls offers a potential solution⁸³, and, in addition, allows one to test not only for association but also for linkage disequilibrium, which should always be present if the polymorphism being tested does actually cause obesity⁸⁴. In studies that examine a single candidate gene, there are no statistical problems arising from simultaneous testing of different loci, but the problem of different phenotypes still remains. This creates a quandary, because testing multiple, partly independent phenotypes raises the threshold required to achieve statistical significance but also offers the potential of increased power by focusing on a subgroup of patients in which the underlying aetiologies are likely to be less heterogeneous than the entire population.

An additional quandary is inherent to the nature of obesity as a complex disease, and, as a consequence, the long list of plausible candidates for study. Ideally, prioritizing this list should be based on comprehensive biochemical, metabolic and model-organism studies⁸⁵; in practice, such data are rarely available and the choice of which candidates to test is made empirically.

Some of the problems discussed above are exemplified by a common polymorphism, W64R (a tryptophan-to-arginine substitution at codon 64), in the β_3 -adrenergic receptor, a catecholamine receptor whose primary function is to activate brown fat thermogenesis in

mice (ref. 86; in addition, see review by Lowell and Spiegelman, pp. 652–660). At first glance, the hypothesis that the W64R variant may impair energy expenditure and cause obesity seems reasonable. However, brown fat in humans occurs mainly in neonates, and the biochemical evidence that the W64R mutation impairs receptor signalling is mixed^{87,88}. In retrospect, then, one can understand why over 40 association and linkage analyses, involving more than 7,000 subjects, have led to markedly inconsistent findings¹⁹. Prospects for resolving this dilemma are not encouraging as the conclusions of two recent meta-analyses also differ with respect to the significance of the association^{89,90}.

What role, then, should candidate genes have in the genetics of common human obesity? Experience gathered from the MC4R may be helpful. With some hindsight, MC4R is an outstanding candidate because pharmacological and genetic studies demonstrated (1) that MC4R function was necessary to prevent adiposity; (2) that a complete loss-of-function caused a large phenotypic effect; and (3) that loss-of-function did not compromise systems that are crucial for development or viability^{46–48,91}. Of course, many genes involved in common human obesity may not meet these criteria. In particular, complete loss-of-function in some genes may have pleiotropic effects, but partial loss-of-function — perhaps the most likely form of polymorphic variation — may affect only body weight (for example, the POMC gene). However, a knockout that causes a large phenotypic effect probably identifies a better candidate than a knockout that causes little or no effect. Given these considerations, what may be most helpful in the short term is not a long list of plausible candidates, but a shorter list of good ones, at least for conventional studies that examine a small number of genes. Applying this approach to a “carefully chosen collection of 37 candidate genes”, Urbanek *et al.*⁹² recently found evidence implicating follistatin in polycystic ovary syndrome. From this perspective, association studies are likely to remain an important tool in quantitative genetics. In fact, in the long term, whole-genome association studies based on SNPs in every gene offer the most powerful approach for detecting genes of small effect^{20,93,94}.

Genes and adiposity in the new millennium

A major goal of identifying genes responsible for common human obesity is to provide more rational approaches to therapy, either by elucidating the underlying pathophysiology or by stratifying patients into groups in which the effectiveness of different treatments can be determined empirically. The success of this approach is just beginning to be realized; although there are now several convincing common obesity loci, there are, as yet, no common obesity genes. Nonetheless, progress is being made rapidly, and new strategies for data analysis^{95,96} coupled with advances in genomic technology⁹⁷ are cause for optimism.

By contrast, the search for monogenic causes of obesity is driven primarily by the goal of understanding basic physiological pathways. At first glance, this search might appear complete. However, previously existing mouse mutations represent only a small fraction of the genome, and new obesity genes are likely to emerge in the next few years as the product of large-scale screens based on chemical mutagenesis of whole animals^{98,99} or transposon insertions in embryonic stem cells^{100,101}. These efforts will probably identify new pathways as well as new components of previously known pathways and, in doing so, provide new targets for therapy. Screens can also be designed to target particular types of mutations, such as those that suppress melanocortin antagonism^{102,103}, or those caused by primary disorders of metabolic rate or nutrient partitioning. The latter point is especially important as our genetic framework for understanding body-weight regulation resides primarily at the level of appetite and energy intake.

Although the quantitative genetic and mendelian approaches differ in primary goals, they may yet converge in both substance and application. New mutations in mouse obesity will provide material

for candidate gene analyses, and the fraction of common human obesity genes that cause monogenic obesity when knocked out in mice could be substantial. Regardless of whether polygenic culprits turn out to be monogenic suspects, the approaches may intersect at the therapeutic level. One might imagine, for example, an array of therapies based on different single-gene disorders, each of which can be tested empirically on subsets of obese patients identified by allele sharing. Thus, genetically based treatments for common human obesity could begin while efforts continue towards molecular identification of the underlying genes. It would be surprising if the rate of progress made this past decade is not outpaced in the next one. □

1. Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* (Addison-Wesley, Harlow, 1995).
2. Bouchard, C. *et al.* The response to long-term overfeeding in identical twins. *N. Engl. J. Med.* **322**, 1477–1482 (1990).
3. Stunkard, A. J. *et al.* An adoption study of human obesity. *N. Engl. J. Med.* **314**, 193–198 (1986).
4. Allison, D. B. *et al.* The heritability of body mass index among an international sample of monozygotic twins reared apart. *Int. J. Obes. Relat. Metab. Disord.* **20**, 501–506 (1996).
5. Comuzzie, A. G. *et al.* Genetic and environmental correlations among hormone levels and measures of body fat accumulation and topography. *J. Clin. Endocrinol. Metab.* **81**, 597–600 (1996).
6. Price, R. A. & Gottesman, I. I. Body fat in identical twins reared apart: roles for genes and environment. *Behav. Genet.* **21**, 1–7 (1991).
7. Maes, H. H., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* **27**, 325–351 (1997).
8. Hill, J. O. & Peters, J. C. Environmental contributions to the obesity epidemic. *Science* **280**, 1371–1374 (1998).
9. World Health Organization. *Obesity: Preventing and Managing the Global Epidemic* (World Health Organization, Geneva, 1998).
10. West, D. B., Waguespack, J. & McCollister, S. Dietary obesity in the mouse: interaction of strain with diet composition. *Am. J. Physiol.* **268**, R658–R665 (1995).
11. Comuzzie, A. G. & Allison, D. B. The search for human obesity genes. *Science* **280**, 1374–1377 (1998).
12. Kissebah, A. H. & Krakower, G. R. Regional adiposity and morbidity. *Physiol. Rev.* **74**, 761–811 (1994).
13. Frankel, W. N. & Schork, N. J. Who's afraid of epistasis? *Nature Genet.* **14**, 371–373 (1996).
14. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* **11**, 241–247 (1995).
15. Comuzzie, A. G. *et al.* A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nature Genet.* **15**, 273–276 (1997).
16. Hager, J. *et al.* A genome-wide scan for human obesity genes reveals a major susceptibility locus on chromosome 10. *Nature Genet.* **20**, 304–308 (1998).
17. Rotimi, C. N. *et al.* The quantitative trait locus on chromosome 2 for serum leptin levels is confirmed in African-Americans. *Diabetes* **48**, 643–644 (1999).
18. Hixson, J. E. *et al.* Normal variation in leptin levels in associated with polymorphisms in the proopiomelanocortin gene, POMC. *J. Clin. Endocrinol. Metab.* **84**, 3187–3191 (1999).
19. Perusse, L., Chagnon, Y. C., Weisnagel, J. & Bouchard, C. The human obesity gene map: the 1998 update. *Obes. Res.* **7**, 111–129 (1999).
20. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
21. Fislir, J. S. & Warden, C. H. Mapping of mouse obesity genes: a generic approach to a complex trait. *J. Nutr.* **127**, 1909S–1916S (1997).
22. West, D. B. Genetics of obesity in humans and animal models. *Endocrinol. Metab. Clin. North Am.* **25**, 801–813 (1996).
23. Pomp, D. & Nielsen, M. K. Quantitative genetics of energy balance — lessons from animal models. *Obes. Res.* **7**, 106–110 (1999).
24. Brockmann, G. A., Haley, C. S., Renne, U., Knott, S. A. & Schwerin, M. Quantitative trait loci affecting body weight and fatness from a mouse line selected for extreme high growth. *Genetics* **150**, 369–381 (1998).
25. Moody, D. E., Pomp, D., Nielsen, M. K. & Van Vleck, L. D. Identification of quantitative trait loci influencing traits related to energy balance in selection and inbred lines of mice. *Genetics* **152**, 699–711 (1999).
26. Goodale, H. A study of the inheritance of body weight in the albino mouse by selection. *J. Hered.* **29**, 101–112 (1938).
27. MacArthur, J. Genetics of body size and related characters. I. Selection of small and large races of the laboratory mouse. *Am. Nat.* **78**, 142–157 (1944).
28. Cheverud, J. M. & Routman, E. J. Epistasis and its contribution to genetic variance components. *Genetics* **139**, 1455–1461 (1995).
29. Cheverud, J. M. *et al.* Quantitative trait loci for murine growth. *Genetics* **142**, 1305–1319 (1996).
30. West, D. B., Goudy-Lefevre, J., York, B. & Truett, G. E. Dietary obesity linked to genetic loci on chromosomes 9 and 15 in a polygenic mouse model. *J. Clin. Invest.* **94**, 1410–1416 (1994).
31. York, B. *et al.* Gene-environment interaction: a significant diet-dependent obesity locus demonstrated in a congenic segment on mouse chromosome 7. *Mamm. Genome* **10**, 457–462 (1999).
32. West, D. B., Waguespack, J., York, B., Goudy-Lefevre, J. & Price, R. A. Genetics of dietary obesity in AKR/J x SWR/J mice: segregation of the trait and identification of a linked locus on chromosome 4. *Mamm. Genome* **5**, 546–552 (1994).
33. Chagnon, Y. C. & Bouchard, C. Genetics of obesity: advances from rodent studies. *Trends Genet.* **12**, 441–444 (1996).
34. Leibel, R. L., Chung, W. K. & Chua, S. C. Jr The molecular genetics of rodent single gene obesities. *J. Biol. Chem.* **272**, 31937–31940 (1997).
35. Coleman, D. L. Effects of parabiosis of obese with diabetes and normal mice. *Diabetologia* **9**, 294–298 (1973).
36. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
37. Friedman, J. M. & Halaas, J. L. Leptin and the regulation of body weight in mammals. *Nature* **395**, 763–770 (1998).
38. Ahima, R. S. *et al.* Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252 (1996).
39. Woods, S. C., Seeley, R. J., Porte, D. Jr & Schwartz, M. W. Signals that regulate food intake and energy homeostasis. *Science* **280**, 1378–1383 (1998).
40. Ahima, R. S., Kelly, J., Elmquist, J. K. & Flier, J. S. Distinct physiologic and neuronal responses to decreased leptin and mild hyperleptinemia. *Endocrinology* **140**, 4923–4931 (1999).
41. Siracusa, L. D. The agouti gene: turned on to yellow. *Trends Genet.* **10**, 423–428 (1994).
42. Michaud, E. J. *et al.* A molecular model for the genetic and phenotypic characteristics of the mouse lethal yellow (Ay) mutation. *Proc. Natl. Acad. Sci. USA* **91**, 2562–2566 (1994).
43. Rossi, M. *et al.* A C-terminal fragment of Agouti-related protein increases feeding and antagonizes the effect of alpha-melanocyte stimulating hormone in vivo. *Endocrinology* **139**, 4428–4431 (1998).
44. Wilson, B. D., Ollmann, M. M. & Barsh, G. S. The role of agouti-related protein in regulating body weight. *Mol. Med. Today* **5**, 250–256 (1999).
45. Shutter, J. R. *et al.* Hypothalamic expression of ART, a novel gene related to agouti, is up-regulated in obese and diabetic mutant mice. *Genes Dev.* **11**, 593–602 (1997).
46. Graham, M., Shutter, J. R., Sarmiento, U., Sarosi, I. & Stark, K. L. Overexpression of Agrp leads to obesity in transgenic mice. *Nature Genet.* **17**, 273–274 (1997).
47. Ollmann, M. M. *et al.* Antagonism of central melanocortin receptors in vitro and in vivo by agouti-related protein. *Science* **278**, 135–138 (1997).
48. Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131–141 (1997).
49. Elmquist, J. K., Elias, C. F. & Saper, C. B. From lesions to leptin: hypothalamic control of food intake and body weight. *Neuron* **22**, 221–232 (1999).
50. Elias, C. F. *et al.* Leptin differentially regulates NPY and POMC neurons projecting to the lateral hypothalamic area. *Neuron* **23**, 775–786 (1999).
51. Mizuno, T. M. & Mobbs, C. V. Hypothalamic agouti-related protein messenger ribonucleic acid is inhibited by leptin and stimulated by fasting. *Endocrinology* **140**, 814–817 (1999).
52. Broberger, C., Johansen, J., Johansson, C., Schalling, M. & Hokfelt, T. The neuropeptide Y/agouti gene-related protein (AGRP) brain circuitry in normal, anorectic, and monosodium glutamate-treated mice. *Proc. Natl. Acad. Sci. USA* **95**, 15043–15048 (1998).
53. Elmquist, J. K., Maratos-Flier, E., Saper, C. B. & Flier, J. S. Unraveling the central nervous system pathways underlying responses to leptin. *Nature Neurosci.* **1**, 445–450 (1998).
54. Seeley, R. *et al.* Melanocortin receptors in leptin effects. *Nature* **390**, 349 (1997).
55. Cheung, C. C., Clifton, D. K. & Steiner, R. A. Proopiomelanocortin neurons are direct targets for leptin in the hypothalamus. *Endocrinology* **138**, 4489–4492 (1997).
56. Marsh, D. J. *et al.* Response of melanocortin-4 receptor-deficient mice to anorectic and orexigenic peptides. *Nature Genet.* **21**, 119–122 (1999).
57. Boston, B., Blyden, K., Varnerin, J. & Cone, R. Independent and additive effects of central POMC and leptin pathways on murine obesity. *Science* **278**, 1641–1644 (1997).
58. Yaswen, L., Diehl, N., Brennan, M. B. & Hochgeschwender, U. Obesity in the mouse model of proopiomelanocortin deficiency responds to peripheral melanocortin. *Nature Med.* **5**, 1066–1070 (1999).
59. Tecott, L. H. *et al.* Eating disorder and epilepsy in mice lacking 5-HT_{2C} serotonin receptors. *Nature* **374**, 542–546 (1995).
60. Hahm, S. *et al.* Targeted deletion of the Vgf gene indicates that the encoded secretory peptide precursor plays a novel role in the regulation of energy balance. *Neuron* **23**, 537–548 (1999).
61. Shimada, M., Tritos, N. A., Lowell, B. B., Flier, J. S. & Maratos-Flier, E. Mice lacking melanin-concentrating hormone are hypophagic and lean. *Nature* **396**, 670–674 (1998).
62. Ohta, T. *et al.* Imprinting-mutation mechanisms in Prader-Willi syndrome. *Am. J. Hum. Genet.* **64**, 397–413 (1999).
63. Kolehmainen, J. *et al.* Refined mapping of the Cohen syndrome gene by linkage disequilibrium. *Eur. J. Hum. Genet.* **5**, 206–213 (1997).
64. Russell-Eggitt, I. M. *et al.* Alstrom syndrome. Report of 22 cases and literature review. *Ophthalmology* **105**, 1274–1280 (1998).
65. Beales, P. L., Warner, A. M., Hitman, G. A., Thakker, R. & Flint, F. A. Bardet-Biedl syndrome: a molecular and phenotypic study of 18 families. *J. Med. Genet.* **34**, 92–98 (1997).
66. Bruford, E. A. *et al.* Linkage mapping in 29 Bardet-Biedl syndrome families confirms loci in chromosomal regions 11q13, 15q22.3–q23, and 16q21. *Genomics* **41**, 93–99 (1997).
67. Mathews, K. D. *et al.* Linkage localization of Borjeson-Forssman-Lehmann syndrome. *Am. J. Med. Genet.* **34**, 470–474 (1989).
68. Gunay-Aygun, M., Cassidy, S. B. & Nicholls, R. D. Prader-Willi and other syndromes associated with obesity and mental retardation. *Behav. Genet.* **27**, 307–324 (1997).
69. Strobel, A., Issad, T., Camoin, L., Ozata, M. & Strosberg, A. D. A leptin missense mutation associated with hypogonadism and morbid obesity. *Nature Genet.* **18**, 213–215 (1998).
70. Montague, C. T. *et al.* Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* **387**, 903–908 (1997).
71. Clement, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398–401 (1998).
72. Bray, G. A. & York, D. A. Genetically transmitted obesity in rodents. *Physiol. Rev.* **51**, 598–646 (1971).
73. Krude, H. *et al.* Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nature Genet.* **19**, 155–157 (1998).
74. Yeo, G. S. *et al.* A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nature Genet.* **20**, 111–112 (1998).
75. Hinney, A. *et al.* Several mutations in the melanocortin-4 receptor gene including a nonsense and a frameshift mutation associated with dominantly inherited obesity in humans. *J. Clin. Endocrinol. Metab.* **84**, 1483–1486 (1999).
76. Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nature Genet.* **20**, 113–114 (1998).
77. Gu, W. *et al.* Identification and functional analysis of novel human melanocortin-4 receptor variants. *Diabetes* **48**, 635–639 (1999).
78. Ho, G. & MacKenzie, R. G. Functional characterization of mutations in melanocortin-4 receptor associated with human obesity. *J. Biol. Chem.* **274**, 35816–35822 (1999).
79. Sina, M. *et al.* Phenotypes in three pedigrees with autosomal dominant obesity caused by haploinsufficiency mutations in the melanocortin-4 receptor gene. *Am. J. Hum. Genet.* **65**, 425–432 (1994).

- 1501–1507 (1999).
80. Farooqi, I. S. *et al.* Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *N. Engl. J. Med.* **341**, 879–884 (1999).
81. Naggert, J. K. *et al.* Hyperproinsulinaemia in obese fat/fat mice associated with a carboxypeptidase E mutation which reduces enzyme activity. *Nature Genet.* **10**, 135–142 (1995).
82. Jackson, R. S. *et al.* Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nature Genet.* **16**, 303–306 (1997).
83. Thomson, G. Mapping disease genes: family-based association studies. *Am. J. Hum. Genet.* **57**, 487–498 (1995).
84. Spielman, R. S. & Ewens, W. J. The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.* **59**, 983–989 (1996).
85. Odunsi, K. & Kidd, K. K. A paradigm for finding genes for a complex human trait: polycystic ovary syndrome and follistatin. *Proc. Natl Acad. Sci. USA* **96**, 8315–8317 (1999).
86. Lowell, B. B. & Flier, J. S. Brown adipose tissue, beta 3-adrenergic receptors, and obesity. *Annu. Rev. Med.* **48**, 307–316 (1997).
87. Pietri-Rouxel, F., St John Manning, B., Gros, J. & Strosberg, A. D. The biochemical effect of the naturally occurring Trp64→Arg mutation on human beta3-adrenoceptor activity. *Eur. J. Biochem.* **247**, 1174–1179 (1997).
88. Li, L. S., Lonnqvist, F., Luthman, H. & Arner, P. Phenotypic characterization of the Trp64Arg polymorphism in the beta3-adrenergic receptor gene in normal weight and obese subjects. *Diabetologia* **39**, 857–860 (1996).
89. Allison, D. B., Heo, M., Faith, M. S. & Pietrobello, A. Meta-analysis of the association of the Trp64Arg polymorphism in the beta3-adrenergic receptor with body mass index. *Int. J. Obes. Relat. Metab. Disord.* **22**, 559–566 (1998).
90. Fujisawa, T., Ikegami, H., Kawaguchi, Y. & Ogihara, T. Meta-analysis of the association of Trp64Arg polymorphism of beta3-adrenergic receptor gene with body mass index. *J. Clin. Endocrinol. Metab.* **83**, 2441–2444 (1998).
91. Fan, W., Boston, B. A., Kesterson, R. A., Hrubby, V. J. & Cone, R. D. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* **385**, 165–168 (1997).
92. Urbanek, M. *et al.* Thirty-seven candidate genes for polycystic ovary syndrome: strongest evidence for linkage is with follistatin. *Proc. Natl Acad. Sci. USA* **96**, 8573–8578 (1999).
93. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genet.* **22**, 139–144 (1999).
94. Long, A. D. & Langley, C. H. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res.* **9**, 720–731 (1999).
95. Allison, D. B. *et al.* Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *Am. J. Hum. Genet.* **65**, 531–544 (1999).
96. Allison, D. B. & Schork, N. J. Selected methodological issues in meiotic mapping of obesity genes in humans: issues of power and efficiency. *Behav. Genet.* **27**, 401–421 (1997).
97. Marth, G. T. *et al.* A general approach to single-nucleotide polymorphism discovery. *Nature Genet.* **23**, 452–456 (1999).
98. Hrabe de Angelis, M. & Balling, R. Large scale ENU screens in the mouse: genetics meets genomics. *Mutat. Res.* **400**, 25–32 (1998).
99. Brown, S. D. & Nolan, P. M. Mouse mutagenesis-systematic studies of mammalian gene function. *Hum. Mol. Genet.* **7**, 1627–1633 (1998).
100. Zambrowicz, B. P. *et al.* Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**, 608–611 (1998).
101. Brennan, J. & Skarnes, W. C. Gene trapping in mouse embryonic stem cells. *Methods Mol. Biol.* **97**, 123–138 (1999).
102. Miller, K. A. *et al.* Genetic studies of the mouse mutations mahogany and mahoganoid. *Genetics* **146**, 1407–1415 (1997).
103. Gunn, T. M. *et al.* The mouse mahogany locus encodes a transmembrane form of human attractin. *Nature* **398**, 152–156 (1999).
104. Norman, R. A. *et al.* Genomewide search for genes influencing percent body fat in Pima Indians: suggestive linkage at chromosome 11q21–q22. Pima Diabetes Gene Group. *Am. J. Hum. Genet.* **60**, 166–173 (1997).
105. Hanson, R. L. *et al.* An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am. J. Hum. Genet.* **63**, 1130–1138 (1998).
106. Lee, J. H. *et al.* Genome scan for human obesity and linkage to markers in 20q13. *Am. J. Hum. Genet.* **64**, 196–209 (1999).
107. Collins, A. C., Martin, I. C. & Kirkpatrick, B. W. Growth quantitative trait loci (QTL) on mouse chromosome 10 in a Quackenbush-Swiss x C57BL/6J backcross. *Mamm. Genome* **4**, 454–458 (1993).
108. Warden, C. H. *et al.* Identification of four chromosomal loci determining obesity in a multifactorial mouse model. *J. Clin. Invest.* **95**, 1545–1552 (1995).
109. Dragani, T. A. *et al.* Mapping of body weight loci on mouse chromosome X. *Mamm. Genome* **6**, 778–781 (1995).
110. Taylor, B. A. & Phillips, S. J. Detection of obesity QTLs on mouse chromosomes 1 and 7 by selective DNA pooling. *Genomics* **34**, 389–398 (1996).
111. York, B., Lei, K. & West, D. B. Sensitivity to dietary obesity linked to a locus on chromosome 15 in a CAST/Ei x C57BL/6J F2 intercross. *Mamm. Genome* **7**, 677–681 (1996).
112. Keightley, P. D., Hardge, T., May, L. & Bulfield, G. A genetic map of quantitative trait loci for body weight in the mouse. *Genetics* **142**, 227–235 (1996).
113. Taylor, B. A. & Phillips, S. J. Obesity QTLs on mouse chromosomes 2 and 17. *Genomics* **43**, 249–257 (1997).
114. Lemberas, A. V. *et al.* Identification of an obesity quantitative trait locus on mouse chromosome 2 and evidence of linkage to body fat and insulin on the human homologous region 20q. *J. Clin. Invest.* **100**, 1240–1247 (1997).
115. Mehrabian, M., Wen, P. Z., Fisler, J., Davis, R. C. & Lusis, A. J. Genetic loci controlling body fat, lipoprotein metabolism, and insulin levels in a multifactorial mouse model. *J. Clin. Invest.* **101**, 2485–2496 (1998).
116. Suto, J., Matsuura, S., Imamura, K., Yamanaka, H. & Sekikawa, K. Genetics of obesity in KK mouse and effects of A(y) allele on quantitative regulation. *Mamm. Genome* **9**, 506–510 (1998).
117. Taylor, B. A., Tarantino, L. M. & Phillips, S. J. Gender-influenced obesity QTLs identified in a cross involving the KK type II diabetes-prone mouse strain. *Mamm. Genome* **10**, 963–968 (1999).
118. Tartaglia, L. A. *et al.* Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**, 1263–1271 (1995).
119. Chen, H. *et al.* Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in db/db mice. *Cell* **84**, 491–495 (1996).
120. Lee, G. H. *et al.* Abnormal splicing of the leptin receptor in diabetic mice. *Nature* **379**, 632–635 (1996).
121. Miller, M. W. *et al.* Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the Lethal-Yellow mutation. *Genes Dev.* **7**, 454–467 (1993).
122. Bultman, S. J., Michaud, E. J. & Woychik, R. P. Molecular characterization of the mouse agouti locus. *Cell* **71**, 1195–1204 (1992).
123. Good, D. J. *et al.* Hypogonadism and obesity in mice with a targeted deletion of the Nhlh2 gene. *Nature Genet.* **15**, 397–401 (1997).
124. Boggon, T. J., Shan, W. S., Santagata, S., Myers, S. C. & Shapiro, L. Implication of tubby proteins as transcription factors by structure-based functional analysis. *Science* **286**, 2119–2125 (1999).
125. Noben-Trauth, K., Naggert, J. K., North, M. A. & Nishina, P. M. A candidate gene for the mouse mutation tubby. *Nature* **380**, 534–538 (1996).

Acknowledgements

We thank K. Clement, N. Risch and G. Yeo for helpful advice and discussion. G.S.B. is an Associate Investigator of the Howard Hughes Medical Institute. I.S.F. is a Wellcome Trust Clinical Training Fellow.

Towards a molecular understanding of adaptive thermogenesis

Bradford B. Lowell* & Bruce M. Spiegelman†

*Beth Israel Deaconess Medical Center, Harvard Medical School, 99 Brookline Avenue, Boston, Massachusetts 02215, USA
(e-mail: blowell@caregroup.harvard.edu)

†Dana-Farber Cancer Institute, Harvard Medical School, One Jimmy Fund Way, Smith Building 958, Boston, Massachusetts 02115, USA
(e-mail: bruce_spiegelman@dfci.harvard.edu)

Obesity results when energy intake exceeds energy expenditure. Naturally occurring genetic mutations, as well as ablative lesions, have shown that the brain regulates both aspects of energy balance and that abnormalities in energy expenditure contribute to the development of obesity. Energy can be expended by performing work or producing heat (thermogenesis). Adaptive thermogenesis, or the regulated production of heat, is influenced by environmental temperature and diet. Mitochondria, the organelles that convert food to carbon dioxide, water and ATP, are fundamental in mediating effects on energy dissipation. Recently, there have been significant advances in understanding the molecular regulation of energy expenditure in mitochondria and the mechanisms of transcriptional control of mitochondrial genes. Here we explore these developments in relation to classical physiological views of adaptive thermogenesis.

It is useful to analyse energy expenditure from a thermodynamic perspective. Such assessment treats the organism as a black box, with energy entering as food and exiting as heat and work (Fig. 1). Obesity is the result of energy imbalance over time and, owing to its cumulative nature, it can develop when energy intake exceeds energy expenditure by only a small margin. Total body energy expenditure represents the conversion of oxygen and food (or stored forms of energy such as fat, glycogen and protein) to carbon dioxide, water, heat and work on the environment. The generation of heat is due to the fact that many reactions in energy metabolism, such as those catalysed by the mitochondrial respiratory chain, and those that consume ATP (for example, Na^+/K^+ ATPase, Ca^{2+} ATPase, and actinomyosin ATPase), are exothermic in the forward direction. Work performed on the environment by the organism plus heat released during biological combustion of food equals the amount of energy given off as heat, measured as calories, during 'physical combustion' of food. When the organism is at rest, and therefore not performing work on the environment, energy expenditure can be measured directly as heat produced (direct calorimetry), hence the

term thermogenesis, or indirectly as the amount of oxygen consumed (indirect calorimetry) (Box 1).

Adaptive thermogenesis, also referred to as facultative thermogenesis, is defined operationally as heat production in response to environmental temperature or diet, and serves the purpose of protecting the organism from cold exposure or regulating energy balance after changes in diet. The term usually refers to adaptations of an individual but could also be viewed in the context of adaptations by different phyla or even different species within a phylum to organism–environment interactions (Box 1). In rodents, brown adipose tissue is a major site of adaptive thermogenesis (see later).

Factors influencing adaptive thermogenesis

Cold-induced thermogenesis

Energy expenditure at rest changes markedly in response to environmental temperature. Oxygen consumption increases nearly two- to fourfold in rodents after both acute and chronic cold exposure (4°C)^{1,2}. A portion of the acute response is due to shivering. However, with adaptation, shivering disappears^{1,2} and other mechanisms become prominent, including increased adaptive thermogenesis in

Box 1

Standard metabolic rate

Standard metabolic rate (SMR) is the amount of energy expended when an adult organism is awake but resting, not actively digesting food and is at thermoneutrality (an environmental temperature where heat production is not stimulated, $\sim 28^\circ\text{C}$ for adult humans). Because work is not performed on the environment, all energy expended is released as heat. Metabolic rate decreases below SMR by $\sim 10\%$ during sleep and by as much as 40% with starvation, and increases above SMR by as much as 10–20-fold for short periods during vigorous exercise²³. Metabolic rate in sedentary, adult humans tends to be approximately 150–200% of SMR, with the increase above SMR being due to physical activity and the thermic effect of food. SMR varies greatly between phyla, with endotherms having a fivefold greater metabolic rate than ectotherms of the same size²³, indicating a phylogenetic adaptation necessary for the warm-blooded state. In addition, SMR per gram of body weight varies inversely with body size between species within a phylum⁸³, indicating a scaling adaptation necessary for smaller animals which have a larger surface area (site of heat loss) relative to their volume. Indeed, the metabolic rate of a mouse on a per weight basis is approximately ten times greater than that of a horse⁸³.

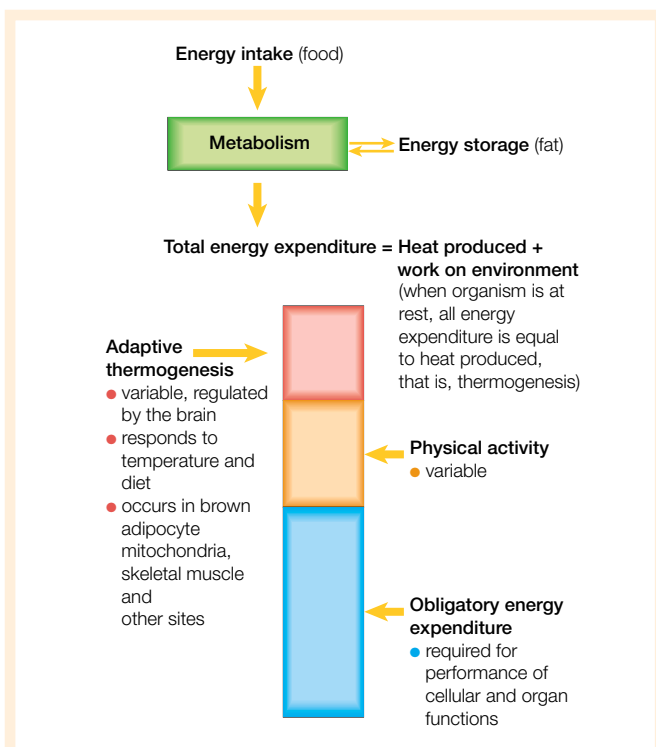


Figure 1 Thermodynamic perspective of energy expenditure. Energy enters an organism as food and exits as heat and work. Energy can also be mobilized from adipose stores. Total energy expenditure can be subdivided into three principal components: obligatory energy expenditure required for normal functioning of cells and organs; energy expenditure resulting from physical activity; and expenditure attributed to adaptive thermogenesis, which is defined as heat production in response to environmental temperature or diet.

brown adipose and possibly other tissues³. Energy expenditure in humans is also sensitive to environmental temperature, but the effect on metabolic rate is smaller. Lowering temperature from 28 to 22 °C has been reported to cause a 7% increase in heat production in identically clothed humans⁴. In general, humans, as opposed to rodents, have a broad thermoneutral zone with relatively small changes in metabolic rate occurring over relatively wide temperature changes. This difference is due, in part, to behavioural responses such as adjustments in the amount of clothing.

Diet-induced thermogenesis

Diet is also a potent regulator of adaptive thermogenesis. Starvation can decrease resting metabolic rate by as much as 40% (ref. 5). Similarly, food restriction sufficient to maintain a 10% reduction in body weight is associated with decreased energy expenditure⁶. The adaptive value of decreasing energy expenditure when food intake is limited is obvious. However, this response is counter-productive during dieting, contributing importantly to the poor long-term efficacy of this treatment for obesity.

Feeding, on the other hand, increases energy expenditure, having both acute and chronic effects on metabolic rate. Feeding acutely increases metabolic rate by ~25–40% in humans and rodents, a phenomenon referred to as the thermic effect of food^{7,8}. Long-term overfeeding also increases energy expenditure⁹. The consequence of increased energy expenditure with overfeeding is a relative protection against the development of obesity. Of interest, this protective adaptation is influenced by genetic background¹⁰, and abnormal responses could contribute to the development of obesity.

Diet-induced thermogenesis is particularly apparent during ingestion of diets that are low in protein. Food serves two important functions: provision of calories to meet energy demands and provision of amino acids to maintain rates of protein synthesis. If the diet is

low in protein, then food intake must be increased to obtain enough protein to sustain protein biosynthesis. This would lead to obesity if the organism lacked the capacity to 'waste' excess calories. Indeed, metabolic efficiency, or the ability to store ingested calories, is decreased by as much as 40% when rodents are fed low-protein diets^{11,12}. This effect may be mediated, at least in part, by stimulation of thermogenesis in brown adipose tissue^{11,12}.

The brain regulates adaptive thermogenesis

Exposure to cold is detected by the brain, leading to activation of efferent pathways controlling energy dissipation. The main effector component of this response is the sympathetic nervous system, which heavily innervates thermogenic targets such as brown adipose tissue and skeletal muscle. Indeed, animals treated with various blockers of the sympathetic nervous system, as well as mice lacking noradrenaline and adrenaline as a result of knockout of the dopamine β -hydroxylase gene, are unable to maintain body temperature during cold exposure^{13,14}. In addition, administration of sympathomimetic agents, such as β -adrenergic-receptor agonists, cause an increase in energy expenditure which is comparable in magnitude to that induced by cold¹³.

Further evidence for central control of adaptive thermogenesis comes from experimental animals with hypothalamic lesions. Destruction of neurons in the hypothalamus either by physical or chemical means results in obesity (reviewed in ref. 15). Typically, the obesity is associated with increased food intake. However, if food intake is restricted so that it equals that observed in controls, obesity still develops¹⁶. This increased storage of calories, during normal caloric intake, indicates that energy expenditure is decreased. These studies show that the brain has the capacity to control adaptive thermogenesis. The presence of the adipocyte-derived hormone leptin and neuropeptides, both of which regulate energy balance in the hypothalamus, is further evidence for regulation of thermogenesis by the brain and is discussed in detail in the review by Schwartz *et al.*, pp. 661–671.

The brain also affects energy expenditure by means of the hypothalamic–pituitary–thyroid axis. It is clear that increases or decreases in thyroid hormone are associated with parallel changes in energy expenditure and that relatively small changes in hormone produce significant effects¹⁷. The mechanism by which thyroid hormone stimulates thermogenesis is not established, but it seems to be due to multiple effects on various aspects of energy metabolism such as substrate cycling, ion cycling and mitochondrial proton leaks¹⁸ (reviewed in ref. 19). Most researchers have viewed thyroid hormone as having, for the most part, a permissive role in adaptive thermogenesis. This is because thyroid hormone levels seem not to be modulated during cold exposure or consumption of high-calorie diets. However, this view may not be entirely correct as thyroid hormone levels have been found to rise in some models of increased caloric intake²⁰ and, importantly, to drop during starvation, an effect which is dependent upon falling leptin levels and mediated by decreased expression of hypothalamic thyrotropin-releasing hormone^{21,22}. This indicates that falling thyroid hormone levels may contribute to starvation-induced decreases in thermogenesis.

Mitochondria convert food to ATP, CO₂ and H₂O

As explained in Box 2 and shown in Figs 2 and 3, fuel metabolism, the electron transport chain, ATP synthesis and ATP use represent coupled reactions in that fixed amounts of reactants produce stoichiometric amounts of products at each step. For example, conversion of glucose to CO₂ generates fixed amounts of NADH and FADH₂, oxidation of NADH and FADH₂ results in a fixed number of protons being pumped across the mitochondrial inner membrane, re-entry of protons by means of ATP synthase generates fixed amounts of ATP, and enzymatic steps performing cellular work use fixed amounts of ATP. For thermogenesis to increase, the degree of 'coupling' at one or more of these sites must change²³. Alternatively,

the consequences of cellular work resulting from reactions using ATP would need to be 'undone' at an increasing rate, in essence wasting ATP as part of a futile cycle. Such reactions completing futile cycles include muscle relaxation (as part of shivering), ion leaks (Na^+ in and K^+ out across the plasma membrane and Ca^{2+} into the cytosol from intracellular stores) and protein degradation, to name a few (Fig. 3).

Prevailing evidence indicates that coupling of most reactions, however, does not change²³. In mammalian cells, reactions that seem to be completely fixed include the amount of NADH and FADH_2 generated by fuel metabolism, the number of protons pumped during NADH and FADH_2 oxidation (with possible exceptions noted below), the number of protons used by ATP synthase to make ATP, and the amount of ATP used to perform cellular work. One firmly established site of uncoupling is the leakage of protons back across the mitochondrial inner membrane, thus bypassing ATP synthase, and converting energy stored within the protonmotive force directly to heat. Mitochondrial proton leaks are a biophysical property of proteolipid bilayers juxtaposed between a strong protonmotive force¹⁸. They are also catalysed by specific inner-membrane proteins such as uncoupling protein (UCP)-1, UCP-2 and UCP-3. These proteins will be discussed in more detail in the following section.

Other sites of 'uncoupling' may also exist, although evidence in support of their role is less compelling. These include decreased proton pumping by cytochrome oxidase (complex IV), mediated by an allosterically regulated heart- and muscle-specific isoform of

cytochrome oxidase subunit VIa²⁴, or increased activity of the glycerol-phosphate shuttle, which competes with the more efficient aspartate-malate shuttle for transfer of NADH generated during glycolysis into the mitochondria for oxidation. The glycerol-phosphate shuttle is less efficient because, unlike the aspartate-malate shuttle, it converts cytosolic NADH to mitochondrial FADH_2 , which, in contrast to NADH, bypasses the first proton pumping site in the electron transport chain (Fig. 2). It is of interest that transgenic mice overexpressing glycerol-3-phosphate dehydrogenase are lean and have increased thermogenesis²⁵. Finally, the contribution of ion and substrate cycles that consume ATP — such as Na^+ , K^+ and Ca^{2+} ion leaks and protein turnover — to adaptive thermogenesis in mammals is presently unknown, but could be significant. The potential importance of Ca^{2+} ion cycling in muscle is evident from examples presented in Box 3.

Regulation of UCP-1 in brown adipose tissue

Cold-induced adaptive thermogenesis

Brown adipose tissue is heavily innervated by sympathetic nerves, and is responsible for a major portion of thermogenesis during cold exposure in rodents. The primary molecule involved in cold-induced thermogenesis in brown fat is UCP-1, a mitochondrial inner-membrane protein that uncouples proton entry from ATP synthesis (Fig. 2)^{26,27}. Indeed, gene-knockout mice lacking UCP-1 have decreased body temperature during cold exposure²⁸.

Two homologues of UCP-1 have been identified. UCP-2 and

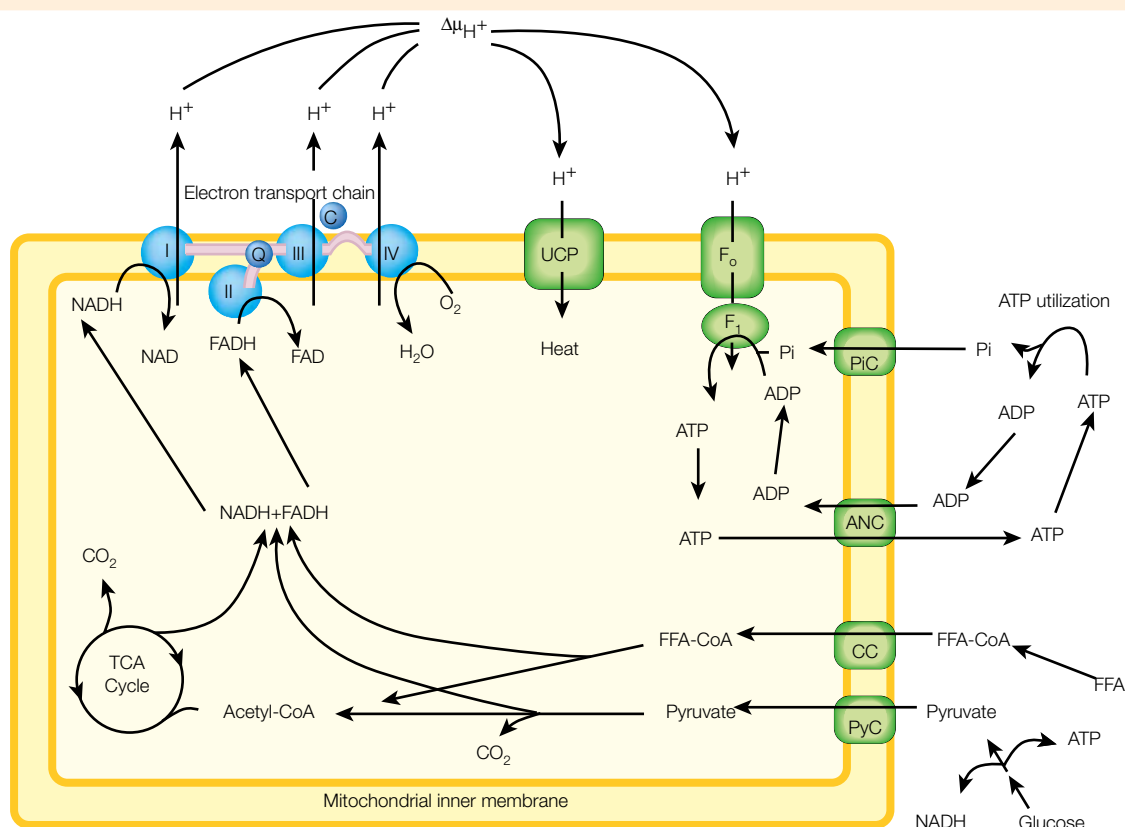


Figure 2 Mitochondrial energy metabolism. Free fatty acids (FFAs) and glucose are oxidized to generate NADH and FADH_2 , which donate electrons to the electron transport chain. Ubiquinone (Q) shuttles electrons from both complexes I and II to complex III, whereas cytochrome *c* (C) shuttles electrons from complex III to complex IV. Molecular oxygen (O_2) is the terminal electron acceptor. Protons are pumped out by complexes I, III and IV of the electron transport chain, which creates a proton electrochemical potential gradient ($\Delta\mu_{\text{H}^+}$). Protons may re-enter the mitochondrial matrix through the F_0F_1 -ATPase, with energy being used to generate ATP from ADP and Pi. Protons may

also re-enter through an uncoupling protein (UCP), with energy being released in the form of heat. Proton re-entry by means of ATP synthase depends upon the availability of ADP, which is generated in the cytosol from reactions using ATP. Abbreviations: ANC, adenine nucleotide carrier; CC, carnitine carrier; complex I, NADH-ubiquinone oxidoreductase; complex II, succinate-ubiquinone oxidoreductase; complex III, ubiquinone-cytochrome-*c* oxidoreductase; complex IV, cytochrome-*c* oxidase; PiC, phosphate carrier; PyC, pyruvate carrier.

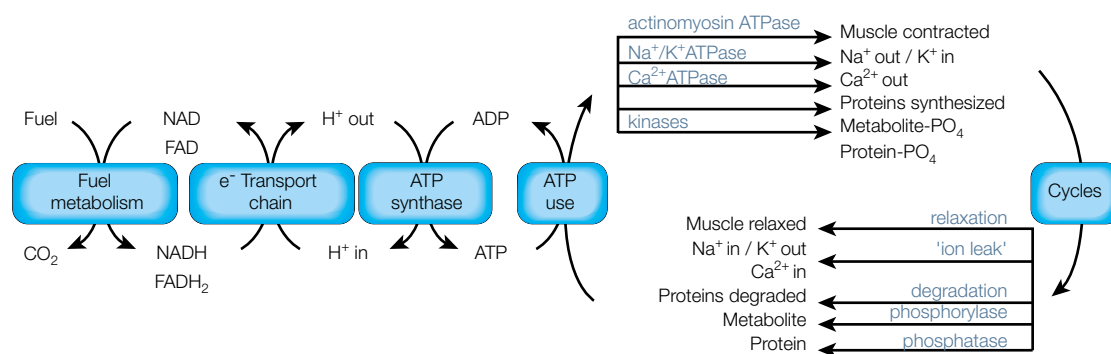


Figure 3 Coupling of reactions in energy metabolism. Metabolism of fuel generates a stoichiometric amount of NADH and FADH₂. Oxidation of NADH and FADH₂ results in ten and six protons, respectively, being pumped out of the mitochondrial matrix. Three protons enter by means of ATP synthase to synthesize one molecule of ATP from ADP and Pi. One additional proton enters the matrix as it is co-transported with Pi through

the phosphate carrier. ATP is then used to perform a fixed amount of work. The major consumers of ATP are shown above. Muscle relaxation, ion leaks, protein degradation and dephosphorylation create the possibility for 'futile cycles'. See ref. 23 for a complete analysis of the concept of coupling with respect to reactions in energy metabolism.

UCP-3 are 73% identical to each other and both are 56% identical to UCP-1 (refs 29–33). UCP-2 is expressed in most tissues at varying levels, whereas UCP-3 is expressed predominantly in skeletal muscle and brown adipose tissue (all three uncoupling proteins are expressed abundantly in brown adipose tissue). Several studies indicate that these UCPs also have proton transport activity^{29,30,33–35}, including those studies using reconstituted proteoliposomes³⁶. In addition, proteins with lower homology to UCP-1 also exist and these too may have proton transport activity^{37,38}. Given that UCP-2 and UCP-3 have uncoupling activity and are expressed widely, it has been hypothesized that they could contribute significantly to adaptive thermogenesis. Arguing against this view, however, are observations that the expression of UCP-2 and UCP-3 messenger RNA increases with starvation^{39,40}, a state known to be associated with decreased energy expenditure. Thus, the function of these UCP homologues, with respect to thermogenesis and regulation of mitochondrial

energy metabolism, is presently uncertain and is an active area of investigation.

β -Adrenergic-receptor stimulation, due to cold exposure or pharmacological agents, has both acute and chronic effects on brown adipose tissue (Fig. 4). UCP-1 activity increases within seconds of stimulation, while chronic stimulation over hours and days results in increased amounts of UCP-1 protein, mitochondrial biogenesis, and both hyperplasia and hypertrophy of brown adipose tissue. The coordination of mitochondrial biogenesis with uncoupling is crucial both in allowing for an increased capacity to generate heat, and in providing the means to maintain appropriate cellular ATP levels in the presence of a mitochondrial proton leak. Acute stimulation of UCP-1 activity is due to increased amounts of cyclic AMP, which activates lipolysis^{41,42}. The resulting increase in free fatty acids is thought to stimulate UCP-1 activity by one of two possible mechanisms. Either fatty acid carboxyl groups function as H⁺ donors to the UCP-1

Box 2

Fuel metabolism and oxidative phosphorylation

Energy is released when food is combusted to carbon dioxide and water. The organism controls this combustion such that energy can be channelled to perform work within the cell. This is accomplished by enzymatically controlled fuel metabolism and mitochondrial oxidative phosphorylation, step-by-step processes in which a portion of the energy content of fuels (fat, carbohydrate and protein) is converted to ATP. Energy stored in the form of ATP is then used to perform biological work within the cell.

Fuel metabolism and oxidative phosphorylation represent a series of reactions that are shown schematically in Fig. 2. The key task of the cell is to match rates of ATP production to rates of ATP consumption. This is made more complex by the fact that sites of ATP production and ATP use are spatially distinct within cells. It was suggested many years ago that ADP, the by-product of ATP use, controls rates of ATP production⁸⁴. This was based upon the observation that addition of ADP to isolated mitochondria stimulates fuel oxidation, mitochondrial oxygen consumption and ATP synthesis. Such regulation provided a mechanism whereby rates of ATP production might be matched to rates of ATP utilization. The molecular mechanism for this regulation was explained by the chemiosmotic hypothesis of Peter Mitchell, to whom the Nobel prize in chemistry was awarded in 1978⁸⁵.

Metabolism of fuel leads to the production of NADH and FADH₂, which in turn donate electrons to the electron transport chain (Fig. 2). As electrons move down through the complexes of the electron transport chain, protons are pumped outside of the mitochondrial inner membrane, creating an electrochemical potential gradient ($\Delta\mu_{H^+}$). Protons then re-enter the mitochondrial matrix through F₀/F₁-ATP synthase in a reaction that is linked tightly to the synthesis of ATP from ADP. If ADP is unavailable, protons are unable to enter through ATP synthase. The three key features of the chemiosmotic hypothesis are as follows: (1) energy derived from fuel oxidation and mitochondrial respiration is converted to $\Delta\mu_{H^+}$; (2) when ADP is available, protons enter via ATP synthase, converting ADP to ATP; and (3) elevated $\Delta\mu_{H^+}$ puts 'backpressure' on proton pumps in the electron transport chain, inhibiting further fuel oxidation. Thus, when ADP is unavailable, fuel oxidation and mitochondrial respiration decreases.

Although the model has many compelling features, several observations indicate that the chemiosmotic hypothesis, when applied to the *in vivo* state, is an oversimplification and that additional layers of regulation must exist. Perhaps the strongest argument comes from observations of increased mitochondrial respiration in the absence of proportional increases in ADP^{86,87}. Such observations have led to the proposal that rates of ATP production increase simultaneously with rates of ATP utilization, possibly as a result of a common activator such as intracellular calcium levels^{88,89}. The mechanism by which such regulation allows precise matching of ATP production and utilization rates, however, is presently unknown.

proton translocation channel²⁷ or UCP-1 transports free fatty acid anions, not protons, from inside to outside of the mitochondrial matrix⁴³. Once outside, the free fatty acids become re-protonated, then flip-flop back across the inner membrane bilayer, creating a protonophore cycle with a net transfer of protons into the mitochondrial matrix.

β_3 -Adrenergic receptors are expressed abundantly and predominantly on brown adipocytes (and also white adipocytes in rodents), and selective agonists of this receptor have been synthesized^{44,45}. Treatment of mice with such agonists doubles oxygen consumption, demonstrating the remarkable capacity of this thermogenic mechanism⁴⁶. In larger mammals, including dogs, cows and primates, discrete deposits of brown fat are present at birth, but become relatively sparse during later development. Chronic treatment with β_3 -adrenergic agonists markedly increases the amount of brown fat in adult dogs⁴⁷ and primates⁴⁸, and brown adipose tissue is abundant in adult humans with catecholamine (that is, adrenaline or noradrenaline)-secreting pheochromocytomas⁴⁹. These data show that a latent source of catecholamine-inducible brown adipocytes exists. Indeed, even in rodents, the efficacy of β_3 -adrenergic agonists to prevent or reverse obesity seems to depend on the ability to expand numbers of brown adipocytes in typical white fat depots^{50,51}, a response influenced by genetic background^{51,52}. Although the precise molecular and cellular basis for recruitment of brown adipocytes induced through stimulation of β -adrenergic receptors is not yet established, it is likely to involve signalling and transcriptional pathways outlined later in this review.

Diet-induced adaptive thermogenesis

There is also evidence that brown adipose tissue is important in diet-induced thermogenesis. Sympathetic nerve activity to brown adipose tissue is reduced in many models of obesity, including leptin-deficient *ob/ob* mice¹⁶. Leptin administration, either centrally or peripherally, increases sympathetic nerve activity to brown fat^{53,54} and, as expected, increases UCP-1 mRNA and protein levels^{55–57}. During starvation, which reduces leptin expression, sympathetic nerve activity to brown fat declines¹³, and this is associated with decreased UCP-1 expression⁵⁸. A role for brown fat in controlling body weight *per se* is illustrated by the fact that transgenic mice with toxigen-mediated reduction of brown fat (UCP-DTA mice) develop obesity⁵⁹. This is in contrast to gene-knockout mice lacking UCP-1, which are cold-sensitive but not obese²⁸. The presence of obesity in UCP-DTA transgenic mice, but not UCP-1-knockout mice, could be due to the existence of alternative thermogenic effectors in brown fat, such as UCP-2 or UCP-3, or possibly an appetite regulatory function of brown fat⁶⁰. Alternatively, UCP-DTA mice could have another, as yet unidentified toxin-

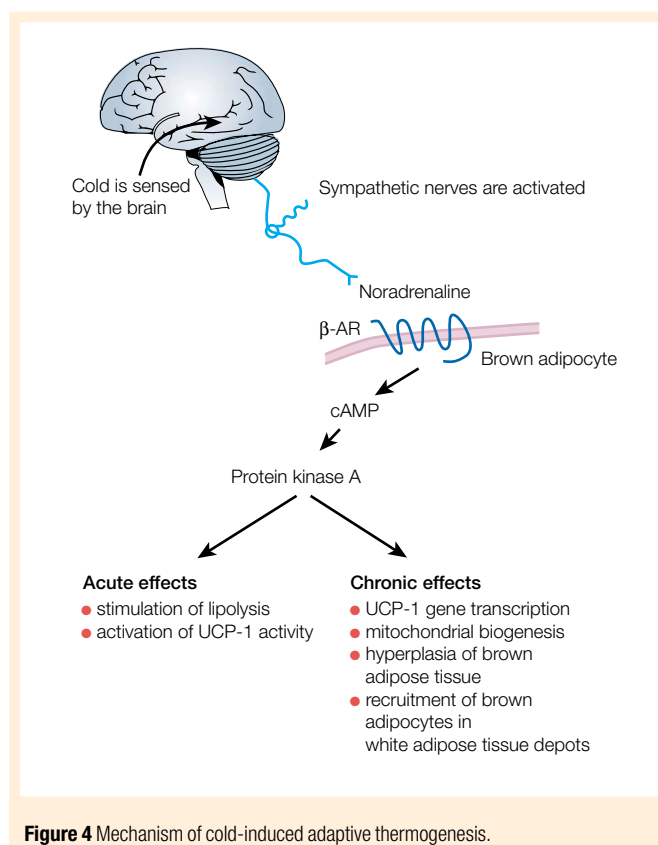


Figure 4 Mechanism of cold-induced adaptive thermogenesis.

induced lesion which causes their obesity. Despite these apparent ambiguities, the evidence for brown fat being important in diet-induced thermogenesis in rodents is strong.

Skeletal muscle as a site for adaptive thermogenesis

Adult humans, unlike rodents, do not have large, distinct depots of brown adipose tissue. But both rodents and humans have varying numbers of brown adipose cells dispersed in white fat depots. Nevertheless, there is considerable suspicion that, in the absence of pharmacological stimulation, brown adipose tissue is important in mediating adaptive thermogenesis in adult humans. Skeletal muscle, on the other hand, represents up to 40% of total body weight and is endowed with significant mitochondrial capacity, causing many researchers to investigate its contribution to adaptive thermogenesis. It has been observed that resting energy expenditure is variable in humans and that low energy expenditure is predictive of future

Box 3

Ca^{2+} ion cycling and thermogenesis

Two examples highlight the potential for effects of ion cycling on energy expenditure: the heater organ of fish and the pathological condition known as malignant hyperthermia.

In certain species of fish, there is a well characterized mechanism for thermogenesis resulting from regulated ion cycling^{90,91}. The 'heater organ' is a derivative of muscle that is relatively devoid of contractile elements. These specialized cells make up most of the superior rectus muscle in the orbit and generate heat for the brain and eyes during cold-water dives. Like typical muscle cells, heater cells possess abundant acetylcholine receptors and have an extensive network of sarcoplasmic reticulum and T-tubules. Mitochondria are also extremely abundant in heater cells, comprising over 60% of total cell volume. Thermogenesis in heater cells is initiated by depolarization, which causes Ca^{2+} release by the sarcoplasmic reticulum. ATP is then consumed by Ca^{2+} -ATPase, which returns Ca^{2+} to the sarcoplasmic reticulum. The increased demand for ATP required to sustain this futile cycle drives fuel oxidation. Thus, depolarization-induced Ca^{2+} entry into the cytoplasm causes ion cycling-mediated thermogenesis.

The thermogenic potential of Ca^{2+} cycling in mammals is evident from the pathological syndrome of malignant hyperthermia, a dominant genetic disorder of humans and pigs which in many cases is due to a mutation in the skeletal muscle ryanodine receptor, the Ca^{2+} -release channel of the sarcoplasmic reticulum⁹². Abnormal Ca^{2+} release, triggered by anaesthesia and/or stress, causes intense thermogenesis, which leads to hyperthermia. Ca^{2+} -ATPase and ryanodine-receptor content increase in the sarcoplasmic reticulum during cold acclimation in birds⁹³, which are notable for their lack of brown adipose tissue, raising the possibility that Ca^{2+} cycling in muscle could contribute to adaptive thermogenesis.

weight gain^{61,62}. A significant portion of the variation in metabolic rate between humans can be accounted for by differences in skeletal muscle energy expenditure⁶³, and further support for a probable role of skeletal muscle in mediating adaptive thermogenesis comes from the demonstration that adrenaline infusion, which causes a 25% increase in whole body energy expenditure in humans, stimulates forearm muscle oxygen consumption by as much as 90% (ref. 64). Assuming that forearm muscle is representative of total body musculature, skeletal muscle then accounts for ~40% of adrenaline infusion-induced thermogenesis in humans. The mechanism for this effect is unknown but could include effects on mitochondrial function and uncoupling, Ca^{2+} cycling, or both. Also unknown is the contribution of other tissues, such as liver and white adipose tissue, to adaptive thermogenesis in humans.

Transcriptional control of mitochondrial genes

A thorough understanding of the transcriptional basis of adaptive thermogenesis must account for the regulation and temporal coordination of mitochondrial biogenesis, expression of uncoupling proteins in a tissue-selective manner, and sensitivity to key hormones such as β -adrenergic agents and thyroid hormone. At best, all of these aspects are only partly understood. As described above, the therapeutic potential for modulation of these systems has provided a strong

incentive to develop a detailed understanding of the relevant gene regulatory programmes.

The transcriptional regulation of gene expression related to mitochondrial proliferation is beginning to be unravelled. Through analysis of the promoters of mitochondrial genes encoded in the cell nucleus, the nuclear respiratory factors (NRF)-1 and -2 have been identified as key components. NRF-1 and -2 bind to and activate the promoters of many genes of the mitochondrial electron transport system such as cytochrome *c*, cytochrome *c* oxidase subunits II and IV, and subunits of the F_0/F_1 -ATP synthase (reviewed in ref. 65). Another target of the NRFs is mitochondrial transcription factor (mtTF)-A, a gene encoded in the nucleus, whose protein product translocates into the mitochondria and stimulates transcription and replication of the mitochondrial genome⁶⁶. Somewhat surprisingly, little quantitative regulation of these components has been demonstrated directly in adaptive thermogenesis *per se*, although altered activity of the ATP synthase promoter through an NRF-2 binding site has been shown during brown fat cell differentiation⁶⁷.

There are many studies that have pointed to thyroid hormone as a major regulator of mitochondrial biogenesis and mitochondrial function *in vivo*. In addition, mitochondrial gene expression is reduced in hypothyroid animals and stimulated upon administration of thyroid hormone. Certain genes of mitochondrial structure

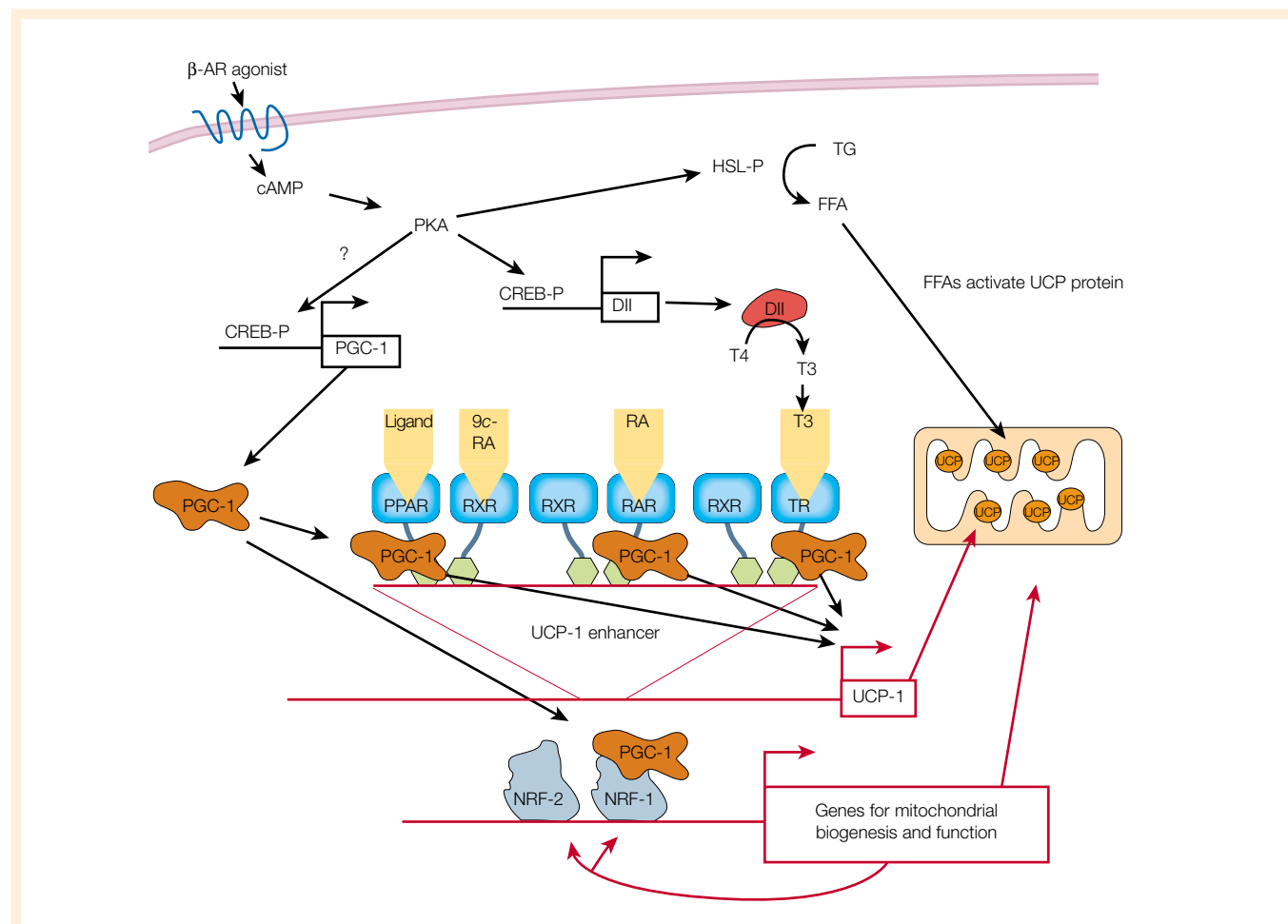


Figure 5 Pathway for β -adrenergic activation of thermogenesis in brown adipocytes. β -adrenergic-receptor (β -AR) agonists stimulate generation of cAMP, which in turn activates protein kinase A (PKA). PKA phosphorylates CREB, which leads to increased gene transcription. It is hypothesized that activated CREB directly induces expression of PGC-1 and the DII. PGC-1 co-activates transcription factors assembled on the UCP-1 enhancer, thus increasing UCP-1 gene expression. In addition, DII increases synthesis of triiodothyronine (T3), the ligand for the thyroid hormone receptor, further

increasing UCP-1 gene expression. PKA also activates hormone-sensitive lipase (HSL), increasing the concentration of free fatty acids (FFAs) which in turn activate UCP-1 protein activity. PGC-1 also co-activates the transcription factor NRF-1, which leads to an increase in genes required for mitochondrial biogenesis, including NRF-1 and NRF-2. This results in marked stimulation of mitochondrial biogenesis. Abbreviations: RXR, retinoid X receptor; RAR, retinoic acid receptor; 9c-RA, 9-*cis*-retinoic acid; RA, retinoic acid; TG, triglyceride.

and function encoded in the cell nucleus have thyroid-hormone response elements, indicating that this hormone is working directly on these genes through thyroid hormone receptors. In addition to these effects on nuclear genes, there have also been several reports suggesting that thyroid hormone and its receptor can translocate into mitochondria to affect transcription patterns⁶⁸; these studies must be considered preliminary until further details become available.

Regulation of UCP-1 gene expression

As a prototypical thermogenic molecule, the promoter of the gene encoding UCP-1 has received extensive attention and analysis. Studies from several laboratories have identified a 220-base-pair enhancer element, located approximately 2.4 kilobases upstream of the mouse and rat UCP-1 genes, which promotes transcription that is both brown fat-selective and responsive to β -adrenergic stimulation (through cAMP)^{69,70}. This complex enhancer element has putative binding sites for the thyroid hormone receptor, retinoic acid receptor and peroxisome proliferator-activated receptor- γ (PPAR- γ), a nuclear receptor expressed in both white and brown fat. UCP-1 in brown fat can be strongly induced by the addition of cAMP, or by constitutive activation of protein kinase A⁷¹. But so far there have been no reports showing a direct role for CREB (cAMP-responsive element binding protein) in the UCP-1 enhancer. Hence, although a role for CREB is possible, or even likely in adaptive thermogenesis and UCP-1 expression, cAMP may work indirectly (see later discussion).

The study of gene expression in brown fat is of special interest because this is the only tissue in the mammalian body that functions exclusively as a thermogenic organ. Although brown fat-selective transcription factors have not been identified, the binding of PPAR- γ is essential for the function of the UCP-1 enhancer⁷². Furthermore, PPAR- γ can activate the UCP-1 enhancer only in brown fat preadipocytes and not other fibroblasts, indicating that molecules interacting with PPAR- γ could represent key components of the selectivity and thermogenic response of brown adipose tissue. A regulatory role for the PPAR- γ system in brown fat development and UCP-1 expression was also shown through administration of PPAR- γ ligands, the synthetic thiazolidinedione (TZD) drugs, to brown fat cells *in vitro* and to mice^{73,74}. These studies showed that activation of PPAR- γ promoted brown fat cell differentiation and UCP-1 expression in culture, and hypertrophy of brown fat tissue *in vivo*.

Although studies of the transcriptional control of the UCP-2 and UCP-3 genes are just beginning, the PPARs also seem to regulate the expression of these proteins. As mentioned above, UCP-2 and UCP-3 are regulated by various nutritional perturbations in a tissue-selective manner. Present insights into the transcriptional regulation of UCP-2 and UCP-3 have come from *in vivo* administration to animals or *in vitro* treatment of cells with PPAR ligands such as TZDs (selective for PPAR- γ) and fibrates (selective for PPAR- α)^{75,76}. These studies have indicated that PPAR- α and PPAR- γ are positive regulators of UCP-2 and UCP-3, with specificity defined by the tissue or cell-type being examined. Given that fatty acids and/or their derivatives are ligands for these receptors, the PPARs may account for much of the nutritional regulation of UCP-2 and UCP-3.

PGC-1: a co-activator linked to thermogenesis

Most biological programmes that have been studied show dominant regulation at the level of DNA-binding transcriptional factors. However, two lines of evidence indicated that a central regulatory component of the adaptive thermogenic programme could be at the level of a transcriptional co-activator. First, the observation that PPAR- γ binding was essential for UCP-1-enhancer functioning⁷² suggested that an important determinant of the brown fat expression of UCP-1 could be a modulator of PPAR- γ function rather than PPAR- γ itself. Second, administration of the TZD ligands for PPAR- γ *in vivo* indicated that this transcriptional system was simultaneously involved in energy storage and energy dissipation through

increased respiration. This paradox suggested that there was tissue-specific regulation of PPAR- γ function.

A potential answer emerged when a PPAR- γ co-activator, PGC-1, was cloned from a brown fat library using a yeast two-hybrid system with PPAR- γ as bait⁷⁷. This factor is highly expressed in brown but not white fat, and is also expressed in heart, kidney, brain and skeletal muscle. Its connection to adaptive thermogenesis was first shown by its marked and rapid induction in brown fat and skeletal muscle upon cold exposure of mice. This cold induction of PGC-1 is largely due to sympathetic nervous system input through β -adrenergic receptors and cAMP action^{77,78}.

In addition to PPAR- γ , PGC-1 also binds to a variety of other nuclear receptors including the retinoic acid and thyroid hormone receptors, both of which positively regulate expression of UCP-1. Interestingly, the docking of PGC-1 to PPAR- γ is not strongly ligand-dependent, whereas its binding to other nuclear receptors, such as the oestrogen receptor, is almost totally ligand-dependent.

Ectopic expression of PGC-1 in cultured cells activates and coordinates multiple aspects of the adaptive thermogenesis programme. Mitochondrial biogenesis is induced, as are many genes of the electron transport system⁷⁷. The expression of uncoupling proteins is also increased in a cell-selective manner⁷⁹. UCP-1 but not UCP-2 or UCP-3 is induced when PGC-1 is introduced into white fat cells, whereas UCP-2 but not UCP-1 or UCP-3 is induced when PGC-1 is expressed in muscle cells. In both fat and muscle cells, these PGC-1-mediated changes in gene expression are reflected in increased respiration, both coupled and uncoupled⁷⁹. These data also indicate that expression of PGC-1 in the adipose lineage could be important in the developmental bifurcation between white and brown fat cells.

PGC-1 also regulates the NRF system

The detailed mechanisms by which PGC-1 induces mitochondrial biogenesis have been examined. Although this co-activator could, in principle, work through a new set of transcription factors, recent data indicate that it is a powerful regulator of the NRF system. PGC-1 expression in muscle cells stimulates a large induction of both NRF-1 and -2 (ref. 79). In addition, PGC-1 binds to and co-activates the action of NRF-1 on the mtTFA promoter, leading to increased mtTFA expression. Presumably, this protein is the direct effector of transcription and replication of the mitochondrial genome. Taken together, these data suggest a direct linkage whereby changes in environmental condition such as temperature or diet can coordinate multiple factors to execute mitochondrial biogenesis and partly uncoupled respiration.

As previously discussed, β -adrenergic-receptor stimulation and the second messenger cAMP potentially increase UCP-1 gene expression in brown fat. Surprisingly, the UCP-1 gene enhancer that mediates this effect lacks a well defined CREB binding site, which indicates that an indirect mechanism may be involved. Although the proximal promoter of UCP-1 does contain CREB binding sites, these may not be relevant as they are located outside of the enhancer that mediates effects of cold exposure and β -adrenergic-receptor stimulation in transgenic mice⁶⁹.

With the discovery of PGC-1, a potential mechanism for catecholamine-induced activation of UCP-1 gene expression, uncoupling activity and mitochondrial biogenesis can be proposed (Fig. 5). In this model, the UCP-1 enhancer binds PPAR- γ , the retinoic acid receptor and the thyroid receptor, each complexed with the retinoid X receptor. β -Adrenergic-receptor stimulation leads to increased expression of PGC-1, mediated potentially by CREB binding to the PGC-1 promoter. Indeed, acute stimulation *in vivo* by either cold exposure or β_3 -adrenergic agonists rapidly increases PGC-1 mRNA levels by more than 50-fold in brown fat tissue^{77,78}. PGC-1, in turn, co-activates each of the three nuclear receptors assembled on the UCP-1 enhancer. PGC-1 induces NRF-1 and NRF-2, and co-activates NRF-1 as well, thereby simultaneously

stimulating mitochondrial biogenesis and respiratory capacity. β -Adrenergic-receptor stimulation also markedly increases expression of type II thyroxine deiodinase (DII)^{80,81}, an effect that is probably mediated directly by CREB⁸². DII in turn generates active ligand for the thyroid hormone receptor. Finally, increased free fatty acids resulting from lipolysis stimulate the activity of UCP-1 protein. In summary, increased protein kinase A activity initiates a cascade of actions, all leading to increased adaptive thermogenesis in brown fat and presumably skeletal muscle. The model presented in Fig. 5 predicts that PGC-1 is crucial in this response. Although the effectors of uncoupling in muscle are less clear, the effects of PGC-1 on muscle-cell respiration suggests an analogous process occurs in this tissue as well.

Future directions

Adaptive thermogenesis is an increasingly attractive target for the development of antiobesity therapies. As the key molecular components become defined, screening for drugs that increase energy dissipation is becoming a more attainable goal. Over the next few years we are likely to learn whether β_3 -adrenergic agents can be developed with the selectivity and potency to increase adaptive thermogenesis in humans. Because of the multifaceted nature of its actions, PGC-1 may also be an interesting target for the development of new therapeutics, although the ease with which compounds that increase the activity of PGC-1 can be developed is completely unknown.

Several key questions still await answers in basic research. What are the physiological roles of UCP-2 and -3? Do they have a protective role against obesity? What are the dominant thermogenic mechanisms in human muscle: uncoupling protein function, Ca^{2+} cycling or as yet undiscovered pathways? What other tissues besides muscle contribute to adaptive thermogenesis in humans? Is PGC-1 involved only in the execution of the oxidative metabolic programme of brown fat and muscle, or is it also involved in the cellular determination of white versus brown fat, and fast twitch (glycolytic) versus slow twitch (oxidative) muscle fibres? Our ability to answer these questions adequately requires a close synthesis of cellular and molecular approaches with whole-animal physiology. □

- Hart, J. S., Heroux, O. & Depocas, F. Cold acclimation and the electromyogram of unanesthetized rats. *J. Appl. Physiol.* **9**, 404–408 (1956).
- Davis, T. R. A., Johnston, D. R., Bell, F. C. & Cremer, B. J. Regulation of shivering and nonshivering heat production during acclimation of rats. *Am. J. Physiol.* **198**, 471–475 (1960).
- Foster, D. O. & Frydman, M. L. Tissue distribution of cold-induced thermogenesis in conscious warm- or cold-acclimated rats reevaluated from changes in tissue blood flow: the dominant role of brown adipose tissue in the replacement of shivering by nonshivering thermogenesis. *Can. J. Physiol. Pharmacol.* **57**, 257–270 (1979).
- Dauncey, M. J. Influence of mild cold on 24 h energy expenditure, resting metabolism and diet-induced thermogenesis. *Br. J. Nutr.* **45**, 257–267 (1981).
- Blaxter, K. *Energy Metabolism in Animals and Man* (Cambridge Univ. Press, Cambridge, 1989).
- Leibel, R. L., Rosenbaum, M. & Hirsch, J. Changes in energy expenditure resulting from altered body weight. *N. Engl. J. Med.* **332**, 621–628 (1995).
- Sims, E. A. & Danforth, E. Jr Expenditure and storage of energy in man. *J. Clin. Invest.* **79**, 1019–1025 (1987).
- Shibata, H. & Bukowiecki, L. J. Regulatory alterations of daily energy expenditure induced by fasting or overfeeding in unrestrained rats. *J. Appl. Physiol.* **63**, 465–470 (1987).
- Levine, J. A., Eberhardt, N. L. & Jensen, M. D. Role of nonexercise activity thermogenesis in resistance to fat gain in humans. *Science* **283**, 212–214 (1999).
- Bouchard, C. et al. The response to long-term overfeeding in identical twins. *N. Engl. J. Med.* **322**, 1477–1482 (1990).
- Kevonian, A. V., Vander Tuig, J. G. & Romsos, D. R. Consumption of a low protein diet increases norepinephrine turnover in brown adipose tissue of adult rats. *J. Nutr.* **114**, 543–549 (1984).
- Rothwell, N. J. & Stock, M. J. Effect of environmental temperature on energy balance and thermogenesis in rats fed normal or low protein diets. *J. Nutr.* **117**, 833–837 (1987).
- Landsberg, L., Saville, M. E. & Young, J. B. Sympathoadrenal system and regulation of thermogenesis. *Am. J. Physiol.* **247**, E181–E189 (1984).
- Thomas, S. A. & Palmiter, R. D. Thermoregulatory and metabolic phenotypes of mice lacking noradrenaline and adrenaline. *Nature* **387**, 94–97 (1997).
- Elmqvist, J. K., Elias, C. F. & Saper, C. B. From lesions to leptin: hypothalamic control of food intake and body weight. *Neuron* **22**, 221–232 (1999).
- Himms-Hagen, J. Brown adipose tissue thermogenesis and obesity. *Prog. Lipid Res.* **28**, 67–115 (1989).
- al-Adani, H., Hoffer, L. J. & Silva, J. E. Resting energy expenditure is sensitive to small dose changes in patients on chronic thyroid hormone replacement. *J. Clin. Endocrinol. Metab.* **82**, 1118–1125 (1997).
- Brand, M. D. et al. The significance and mechanism of mitochondrial proton conductance. *Int. J. Obes. Relat. Metab. Disord.* **23**(Suppl. 6), S4–S11 (1999).
- Silva, J. E. Thyroid hormone control of thermogenesis and energy balance. *Thyroid* **5**, 481–492 (1995).
- Almeida, N. G., Levitsky, D. A. & Strupp, B. Enhanced thermogenesis during recovery from diet-induced weight gain in the rat. *Am. J. Physiol.* **271**, R1380–R1387 (1996).
- Ahima, R. S. et al. Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252 (1996).
- Legradi, G., Emerson, C. H., Ahima, R. S., Flier, J. S. & Lechan, R. M. Leptin prevents fasting-induced suppression of prothymotrophin-releasing hormone messenger ribonucleic acid in neurons of the hypothalamic paraventricular nucleus. *Endocrinology* **138**, 2569–2576 (1997).
- Rolfe, D. F. & Brown, G. C. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol. Rev.* **77**, 731–758 (1997).
- Kadenbach, B. et al. Regulation of energy transduction and electron transfer in cytochrome c oxidase by adenine nucleotides. *J. Bioenerg. Biomembr.* **30**, 25–33 (1998).
- Kozak, L. P., Kozak, U. C. & Clarke, G. T. Abnormal brown and white fat development in transgenic mice overexpressing glycerol 3-phosphate dehydrogenase. *Genes Dev.* **5**, 2256–2264 (1991).
- Nicholls, D. G. & Locke, R. M. Thermogenic mechanisms in brown fat. *Physiol. Rev.* **64**, 1–64 (1984).
- Klingenberg, M. & Huang, S. G. Structure and function of the uncoupling protein from brown adipose tissue. *Biochim. Biophys. Acta* **1415**, 271–296 (1999).
- Enerback, S. et al. Mice lacking mitochondrial uncoupling protein are cold-sensitive but not obese. *Nature* **387**, 90–94 (1997).
- Fleury, C. et al. Uncoupling protein-2: a novel gene linked to obesity and hyperinsulinemia. *Nature Genet.* **15**, 269–272 (1997).
- Gimeno, R. E. et al. Cloning and characterization of an uncoupling protein homolog: a potential molecular mediator of human thermogenesis. *Diabetes* **46**, 900–906 (1997).
- Boss, O. et al. Uncoupling protein-3: a new member of the mitochondrial carrier family with tissue-specific expression. *FEBS Lett.* **408**, 39–42 (1997).
- Vidal-Puig, A., Solanes, G., Gruijic, D., Flier, J. S. & Lowell, B. B. UCP3: an uncoupling protein homologue expressed preferentially and abundantly in skeletal muscle and brown adipose tissue. *Biochem. Biophys. Res. Commun.* **235**, 79–82 (1997).
- Gong, D. W., He, Y., Karas, M. & Reitman, M. Uncoupling protein-3 is a mediator of thermogenesis regulated by thyroid hormone, beta3-adrenergic agonists, and leptin. *J. Biol. Chem.* **272**, 24129–24132 (1997).
- Hinz, W., Faller, B., Gruninger, S., Gazzotti, P. & Chiesi, M. Recombinant human uncoupling protein-3 increases thermogenesis in yeast cells. *FEBS Lett.* **448**, 57–61 (1999).
- Zhang, C. Y., Hagen, T., Mootha, V. K., Sliker, L. J. & Lowell, B. B. Assessment of uncoupling activity of uncoupling protein 3 using a yeast heterologous expression system. *FEBS Lett.* **449**, 129–134 (1999).
- Jaburek, M. et al. Transport function and regulation of mitochondrial uncoupling proteins 2 and 3. *J. Biol. Chem.* **274**, 26003–26007 (1999).
- Sanchis, D. et al. BMCP1, a novel mitochondrial carrier with high expression in the central nervous system of humans and rodents, and respiration uncoupling activity in recombinant yeast. *J. Biol. Chem.* **273**, 34611–34615 (1998).
- Mao, W. et al. UCP4, a novel brain-specific mitochondrial protein that reduces membrane potential in mammalian cells. *FEBS Lett.* **443**, 326–330 (1999).
- Boss, O. et al. Tissue-dependent upregulation of rat uncoupling protein-2 expression in response to fasting or cold. *FEBS Lett.* **412**, 111–114 (1997).
- Weigle, D. S. et al. Elevated free fatty acids induce uncoupling protein 3 expression in muscle: a potential explanation for the effect of fasting. *Diabetes* **47**, 298–302 (1998).
- Prusiner, S. B., Cannon, B., Ching, T. M. & Lindberg, O. Oxidative metabolism in cells isolated from brown adipose tissue. 2. Catecholamine regulated respiratory control. *Eur. J. Biochem.* **7**, 51–57 (1968).
- Bukowiecki, L. J., Folle, N., Lupien, J. & Paradis, A. Metabolic relationships between lipolysis and respiration in rat brown adipocytes. The role of long chain fatty acids as regulators of mitochondrial respiration and feedback inhibitors of lipolysis. *J. Biol. Chem.* **256**, 12840–12848 (1981).
- Jezek, P. et al. Fatty acid cycling mechanism and mitochondrial uncoupling proteins. *Biochim. Biophys. Acta* **1365**, 319–327 (1998).
- Arch, J. R. et al. Atypical β -adrenoceptor on brown adipocytes as target for anti-obesity drugs. *Nature* **309**, 163–165 (1984).
- Strosberg, A. D. & Pietri-Rouxel, F. Function and regulation of the beta 3-adrenoceptor. *Trends Pharmacol. Sci.* **17**, 373–381 (1996).
- Susulic, V. S. et al. Targeted disruption of the beta 3-adrenergic receptor gene. *J. Biol. Chem.* **270**, 29483–29492 (1995).
- Champigny, O. et al. Beta 3-adrenergic receptor stimulation restores message and expression of brown-fat mitochondrial uncoupling protein in adult dogs. *Proc. Natl Acad. Sci. USA* **88**, 10774–10777 (1991).
- Fisher, M. H. et al. A selective human beta3 adrenergic receptor agonist increases metabolic rate in rhesus monkeys. *J. Clin. Invest.* **101**, 2387–2393 (1998).
- Garruti, G. & Ricquier, D. Analysis of uncoupling protein and its mRNA in adipose tissue deposits of adult humans. *Int. J. Obes. Relat. Metab. Disord.* **16**, 383–390 (1992).
- Himms-Hagen, J. et al. Effect of CL-316,243, a thermogenic beta 3-agonist, on energy balance and brown and white adipose tissues in rats. *Am. J. Physiol.* **266**, R1371–R1382 (1994).
- Collins, S., Daniel, K. W., Petro, A. E. & Surwit, R. S. Strain-specific response to beta 3-adrenergic receptor agonist treatment of diet-induced obesity in mice. *Endocrinology* **138**, 405–413 (1997).
- Guerra, C., Koza, R. A., Yamashita, H., Walsh, K. & Kozak, L. P. Emergence of brown adipocytes in white fat is under genetic control. Effects on body weight and adiposity. *J. Clin. Invest.* **102**, 412–420 (1998).
- Collins, S. et al. Role of leptin in fat regulation. *Nature* **380**, 677 (1996).
- Haynes, W. G., Morgan, D. A., Walsh, S. A., Mark, A. L. & Sivitz, W. I. Receptor-mediated regional sympathetic nerve activation by leptin. *J. Clin. Invest.* **100**, 270–278 (1997).
- Scarpace, P. J., Matheny, M., Pollock, B. H. & Tumer, N. Leptin increases uncoupling protein expression and energy expenditure. *Am. J. Physiol.* **273**, E226–E230 (1997).
- Satoh, N. et al. Satiety effect and sympathetic activation of leptin are mediated by hypothalamic melanocortin system. *Neurosci. Lett.* **249**, 107–110 (1998).
- Cusin, I. et al. Chronic central leptin infusion enhances insulin-stimulated glucose metabolism and favors the expression of uncoupling proteins. *Diabetes* **47**, 1014–1019 (1998).
- Champigny, O. & Ricquier, D. Effects of fasting and refeeding on the level of uncoupling protein mRNA in rat brown adipose tissue: evidence for diet-induced and cold-induced responses. *J. Nutr.* **120**, 1730–1736 (1990).
- Lowell, B. B. et al. Development of obesity in transgenic mice after genetic ablation of brown

- adipose tissue. *Nature* **366**, 740–742 (1993).
60. Melnyk, A. & Himms-Hagen, J. Temperature-dependent feeding: lack of role for leptin and defect in brown adipose tissue-ablated obese mice. *Am. J. Physiol.* **274**, R1131–1135 (1998).
61. Ravussin, E. *et al.* Reduced rate of energy expenditure as a risk factor for body-weight gain. *N. Engl. J. Med.* **318**, 467–472 (1988).
62. Roberts, S. B., Savage, J., Coward, W. A., Chew, B. & Lucas, A. Energy expenditure and intake in infants born to lean and overweight mothers. *N. Engl. J. Med.* **318**, 461–466 (1988).
63. Zurlo, F., Larson, K., Bogardus, C. & Ravussin, E. Skeletal muscle metabolism is a major determinant of resting energy expenditure. *J. Clin. Invest.* **86**, 1423–1427 (1990).
64. Simonsen, L., Bulow, J., Madsen, J. & Christensen, N. J. Thermogenic response to epinephrine in the forearm and abdominal subcutaneous adipose tissue. *Am. J. Physiol.* **263**, E850–E855 (1992).
65. Gupteja, S., Virbasius, C. M. & Scarpulla, R. C. Nuclear respiratory factors 1 and 2 utilize similar glutamine-containing clusters of hydrophobic residues to activate transcription. *Mol. Cell. Biol.* **16**, 5708–5716 (1996).
66. Virbasius, J. V. & Scarpulla, R. C. Activation of the human mitochondrial transcription factor A gene by nuclear respiratory factors: a potential regulatory link between nuclear and mitochondrial gene expression in organelle biogenesis. *Proc. Natl. Acad. Sci. USA* **91**, 1309–1313 (1994).
67. Villena, J. A. *et al.* Regulation of mitochondrial biogenesis in brown adipose tissue: nuclear respiratory factor-2/GA-binding protein is responsible for the transcriptional regulation of the gene for the mitochondrial ATP synthase beta subunit. *Biochem. J.* **331**, 121–127 (1998).
68. Demonacos, C. V. *et al.* Mitochondrial genes as sites of primary action of steroid hormones. *Steroids* **61**, 226–232 (1996).
69. Cassard-Doulcier, A. M. *et al.* Tissue-specific and beta-adrenergic regulation of the mitochondrial uncoupling protein gene: control by cis-acting elements in the 5'-flanking region. *Mol. Endocrinol.* **7**, 497–506 (1993).
70. Kozak, U. C. *et al.* An upstream enhancer regulating brown-fat-specific expression of the mitochondrial uncoupling protein gene. *Mol. Cell. Biol.* **14**, 59–67 (1994).
71. Cummings, D. E. *et al.* Genetically lean mice result from targeted disruption of the RII beta subunit of protein kinase A. *Nature* **382**, 622–626 (1996).
72. Sears, I. B., MacGinnitie, M. A., Kovacs, L. G. & Graves, R. A. Differentiation-dependent expression of the brown adipocyte uncoupling protein gene: regulation by peroxisome proliferator-activated receptor gamma. *Mol. Cell. Biol.* **16**, 3410–3419 (1996).
73. Foellmi-Adams, L. A., Wyse, B. M., Herron, D., Nedergaard, J. & Kletzien, R. F. Induction of uncoupling protein in brown adipose tissue. Synergy between norepinephrine and pioglitazone, an insulin-sensitizing agent. *Biochem. Pharmacol.* **52**, 693–701 (1996).
74. Tai, T. A. C. *et al.* Activation of the nuclear receptor peroxisome proliferator-activated receptor gamma promotes brown adipocyte differentiation. *J. Biol. Chem.* **271**, 29909–29914 (1996).
75. Brun, S. *et al.* Activators of peroxisome proliferator-activated receptor-alpha induce the expression of the uncoupling protein-3 gene in skeletal muscle: a potential mechanism for the lipid intake-dependent activation of uncoupling protein-3 gene expression at birth. *Diabetes* **48**, 1217–1222 (1999).
76. Aubert, J. *et al.* Up-regulation of UCP-2 gene expression by PPAR agonists in preadipose and adipose cells. *Biochem. Biophys. Res. Commun.* **238**, 606–611 (1997).
77. Puigserver, P. *et al.* A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* **92**, 829–839 (1998).
78. Boss, O. *et al.* Role of the beta(3)-adrenergic receptor and/or a putative beta(4)-adrenergic receptor on the expression of uncoupling proteins and peroxisome proliferator-activated receptor-gamma coactivator-1. *Biochem. Biophys. Res. Commun.* **261**, 870–876 (1999).
79. Wu, Z. *et al.* Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell* **98**, 115–124 (1999).
80. Hosoi, Y. *et al.* Expression and regulation of type II iodothyronine deiodinase in cultured human skeletal muscle cells. *J. Clin. Endocrinol. Metab.* **84**, 3293–3300 (1999).
81. Encke, D., Ely, M., Heldmaier, G. & Klaus, S. Physiological approach to maturation of brown adipocytes in primary cell culture. *Biochim. Biophys. Acta* **1357**, 339–347 (1997).
82. Bartha, T. *et al.* Characterization of the 5'-flanking and 5'-untranslated regions of the cyclic adenosine 3',5'-monophosphate-responsive human type 2 iodothyronine deiodinase gene. *Endocrinology* **141**, 229–237 (2000).
83. Porter, R. K. & Brand, M. D. Body mass dependence of H⁺ leak in mitochondria and its relevance to metabolic rate. *Nature* **362**, 628–630 (1993).
84. Chance, B. & Williams, C. M. *J. Biol. Chem.* **217**, 405–427 (1955).
85. Mitchell, P. Keilin's respiratory chain concept and its chemiosmotic consequences. *Science* **206**, 1148–1159 (1979).
86. Balaban, R. S. Regulation of oxidative phosphorylation in the mammalian cell. *Am. J. Physiol.* **258**, C377–C389 (1990).
87. Hochachka, P. W. The metabolic implications of intracellular circulation. *Proc. Natl. Acad. Sci. USA* **96**, 12233–12239 (1999).
88. McCormack, J. G., Halestrap, A. P. & Denton, R. M. Role of calcium ions in regulation of mammalian intramitochondrial metabolism. *Physiol. Rev.* **70**, 391–425 (1990).
89. Brown, G. C. Control of respiration and ATP synthesis in mammalian mitochondria and cells. *Biochem. J.* **284**, 1–13 (1992).
90. Block, B. A. Thermogenesis in muscle. *Annu. Rev. Physiol.* **56**, 535–577 (1994).
91. O'Brien, J. & Block, B. A. Effects of Ca²⁺ on oxidative phosphorylation in mitochondria from the thermogenic organ of marlin. *J. Exp. Biol.* **199**, 2679–2687 (1996).
92. Denborough, M. Malignant hyperthermia. *Lancet* **352**, 1131–1136 (1998).
93. Dumonteil, E., Barre, H. & Meissner, G. Expression of sarcoplasmic reticulum Ca²⁺ transport proteins in cold-acclimating ducklings. *Am. J. Physiol.* **269**, C955–C960 (1995).

Central nervous system control of food intake

Michael W. Schwartz*, Stephen C. Woods†, Daniel Porte Jr*, Randy J. Seeley† & Denis G. Baskin*‡

Departments of Medicine* and Biological Structure‡, Harborview Medical Center and VA Puget Sound Health Care System, University of Washington, Seattle, Washington 98104-2499, USA

†Department of Psychiatry, University of Cincinnati, Cincinnati, Ohio 45267-0559, USA

New information regarding neuronal circuits that control food intake and their hormonal regulation has extended our understanding of energy homeostasis, the process whereby energy intake is matched to energy expenditure over time. The profound obesity that results in rodents (and in the rare human case as well) from mutation of key signalling molecules involved in this regulatory system highlights its importance to human health. Although each new signalling pathway discovered in the hypothalamus is a potential target for drug development in the treatment of obesity, the growing number of such signalling molecules indicates that food intake is controlled by a highly complex process. To better understand how energy homeostasis can be achieved, we describe a model that delineates the roles of individual hormonal and neuropeptide signalling pathways in the control of food intake and the means by which obesity can arise from inherited or acquired defects in their function.

For most of us, the composition and amount of food that we eat varies considerably from one meal to the next and from one day to the next. Our common experience, therefore, seems at odds with the hypothesis that food intake is highly regulated. Emotions, social factors, time of day, convenience and cost are but a few of the variables that are not biologically regulated, but nonetheless affect meal-to-meal energy intake. As a consequence, daily energy intake is variable both within and among individuals, and is not well correlated with daily energy expenditure¹. Despite short-term mismatches in energy balance, however, most of us match cumulative energy intake to energy expenditure with great precision when measured over a period that spans many meals¹. This phenomenon reflects an active regulatory process, termed energy homeostasis, that promotes stability in the amount of body energy stored in the form of fat.

Although it is overly simplistic to reduce a behaviour as complex as feeding to a series of molecular interactions, discoveries over the past few years have identified signalling molecules that affect food intake and that are critical for normal energy homeostasis. The application of molecular genetics to mice has been especially important in this effort. For example, several monogenic forms of human obesity were identified by searching for mutations homologous to those causing obesity in mice^{2–5}. Although such monogenic obesity syndromes are rare (see review by Barsh *et al.*, pp. 644–651), the successful use of murine models to study human obesity indicates that substantial homology exists across mammalian species in the functional organization of the weight-regulatory system. More importantly, the identification of molecules that control food intake has generated new targets for drug development in the treatment of obesity and related disorders. Optimism that we may soon enter an era of improved obesity treatment, therefore, seems justified.

Because of the enormous toll on human health taken by obesity and related disorders, an improved understanding of the control of food intake is an important priority. However, the growing number of molecules implicated in energy homeostasis raises nearly limitless possibilities for how body-weight regulation might occur. The aim of this article is to review these advances and to present them in the context of a model for long-term maintenance of energy homeostasis.

Model for energy homeostasis

The increase of food intake (hyperphagia) triggered by a period of fasting is a simple but compelling example of food-intake regulation. The consequent recovery of lost body weight to baseline values, accompanied by the gradual return to normal levels of energy intake⁶, is testimony to a regulatory process that is both precise and robust. To explain this phenomenon, Kennedy proposed⁷ in 1953 that inhibitory signals generated in proportion to body fat stores act in the brain to reduce food intake. Thus, when weight loss induced by caloric restriction reduces the level of these inhibitory signals, food intake increases until the energy deficit is corrected. This model, however, does not explain how energy intake is controlled during individual meals. Twenty years later, Gibbs and Smith proposed⁸ that signals generated during a meal (termed 'satiety factors'), including peptides secreted from the gastrointestinal tract, provide information to the brain that inhibits feeding and leads to meal termination. A model that seems to unite these two hypotheses is illustrated schematically in Fig. 1.

Adiposity signals: leptin and insulin

The pancreatic hormone, insulin, which enters the brain from the circulation⁹ and acts there to reduce energy intake¹⁰, was the first hormonal signal to be implicated in the control of body weight by the central nervous system (CNS). The subsequent demonstration that the profound hyperphagia and obesity of *ob/ob* mice results from autosomal recessive mutation of the gene encoding leptin¹¹, a hormone secreted by adipocytes, provided compelling evidence of a second

Different mechanisms underlie the association of insulin and leptin with body fat content²¹. The effect of weight gain to reduce insulin sensitivity seems to explain how insulin, but not leptin, varies according to body fat stores²². As weight increases, insulin secretion must increase in both the basal state and in response to meals to compensate for insulin resistance if normal glucose homeostasis is to be maintained^{23,24}. Failure of the pancreatic β -cell to achieve this adaptive increase of insulin secretion causes hyperglycaemia, and probably contributes to the association of type 2 diabetes with obesity. Increased insulin secretion as obesity progresses is thus hypothesized to increase insulin delivery to the brain, where it helps to limit further weight gain.

For example, food deprivation lowers plasma leptin concentrations in both rodents and humans much more rapidly and to a greater extent than would be expected from the decrease of body fat content. This exaggerated early decline of leptin levels would enable compensatory responses to be activated before energy stores are substantially depleted.

The diagram illustrates the integrated regulation of energy balance by the Central Nervous System (CNS). It features a central brain icon labeled 'CNS' with two sections: 'Anabolic' (marked with a minus sign) and 'Catabolic' (marked with a plus sign). Four numbered blue boxes provide context for the regulatory pathways:

- Box 1 (Top Right):** Leptin and insulin act on central effector pathways in the hypothalamus, repressing brain anabolic neural circuits that stimulate eating and inhibit energy expenditure, while simultaneously activating catabolic circuits that inhibit food intake and increase energy expenditure.
- Box 2 (Top Left):** Low leptin and insulin levels in the brain during weight loss increase activity of anabolic neural pathways that stimulate eating and suppress energy expenditure, and decrease activity of catabolic pathways that cause anorexia and weight loss.
- Box 3 (Bottom Left):** Ingestion of food generates neural and hormonal satiety signals to the hindbrain. Leptin/insulin-sensitive central effector pathways interact with hindbrain satiety circuits to regulate the meal size, thereby modulating food intake and energy balance.
- Box 4 (Bottom Right):** Leptin and insulin circulate in the blood in concentrations proportional to body fat content and energy balance.

The regulatory pathways are shown with arrows and signs:

- Food intake:** Represented by an icon of a hand eating. It has a green arrow pointing to the 'Anabolic' CNS section (marked '+') and a red arrow pointing to the 'Catabolic' CNS section (marked '-').
- Energy expenditure:** Represented by a campfire icon. It has a green arrow pointing to the 'Anabolic' CNS section (marked '-') and a red arrow pointing to the 'Catabolic' CNS section (marked '+').
- Energy balance:** A yellow box at the bottom center. It receives a green arrow from 'Food intake' (marked '+') and a red arrow from 'Energy expenditure' (marked '-').
- Fat stores:** Represented by a cluster of red adipocytes. It receives a yellow arrow from 'Energy balance' (marked '+').
- Adiposity signals:** A yellow box labeled 'Insulin/leptin' that receives a yellow arrow from 'Fat stores' (marked '+').
- Feedback loops:** Yellow arrows from 'Adiposity signals' point back to the 'Anabolic' and 'Catabolic' sections of the CNS, with a minus sign on the arrow to the 'Anabolic' section and a plus sign on the arrow to the 'Catabolic' section.

inhibiting anabolic, effector pathways. These pathways have opposing effects on energy balance (the difference between calories consumed and energy expended) that in turn determines the amount of body fuel stored as fat.

not insulin, is required for hyperphagia in this model. Thus, although both leptin and insulin probably participate in the CNS control of energy homeostasis, available data indicate that leptin has the more critical role.

Leptin resistance and obesity

The hypothesis that leptin resistance can occur in association with obesity was first suggested by the finding of elevated plasma leptin levels in obese humans¹³. This hypothesis suggests that some cases of human obesity may be due to reduced leptin action in the brain, and

that affected individuals are unlikely to respond to pharmacological treatment with leptin. Resistance to leptin is clearly documented in mice (for example, *db/db*)¹⁹ and rats (for example, *fa/fa*)³¹ bearing mutant leptin receptors, but also in mice that develop obesity for other reasons. These include mice with genetic ablation of thermogenic brown adipose tissue³², mice that lack melanocortin-4 (MC4) receptors³³, agouti (*A^y/a*) mice³⁴ (see later) and mice fed a highly palatable high-fat diet¹⁹.

Several mechanisms may contribute to leptin resistance. By

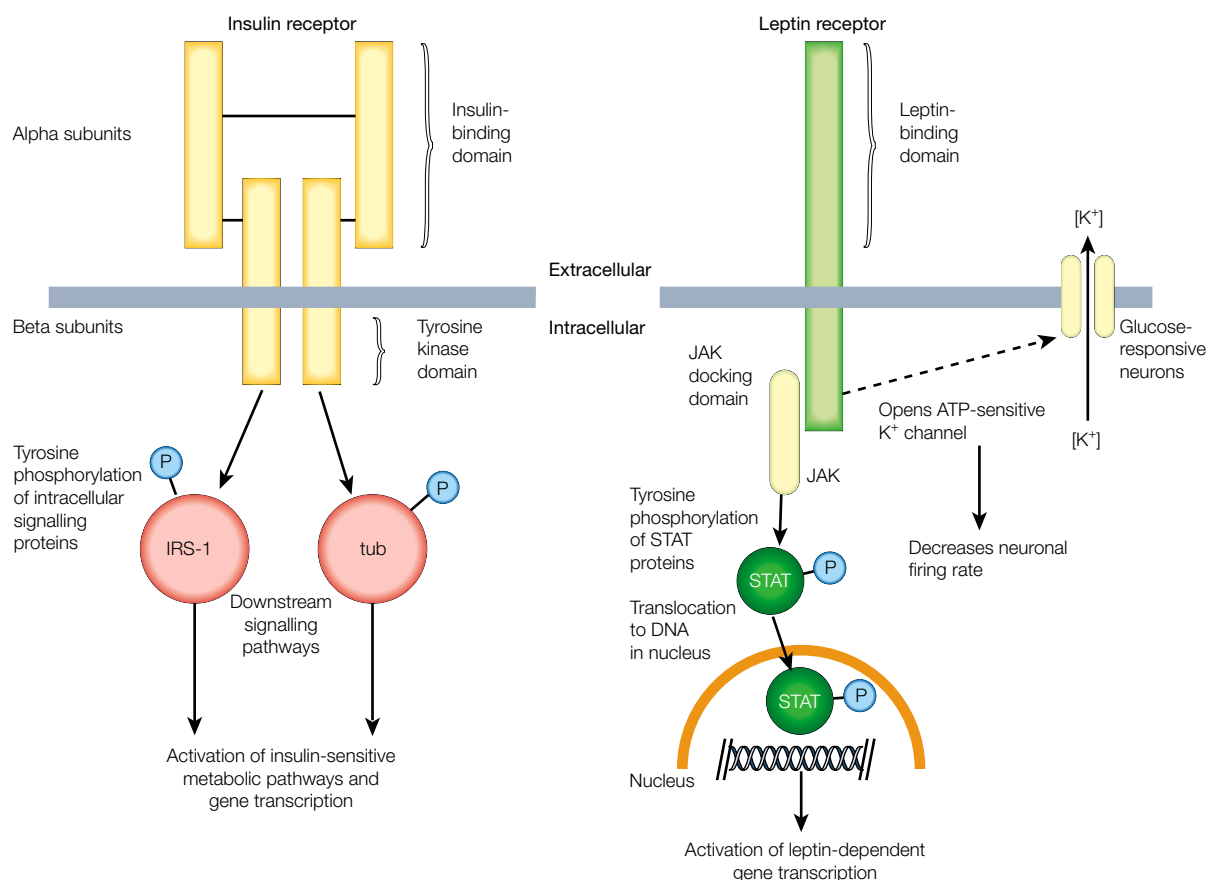
Box 1

Leptin- and insulin-receptor signalling

The insulin receptor comprises an extracellular α -subunit involved in ligand binding and an intracellular β -subunit that transduces the insulin signal to the cell (Box 1 Figure). The β -subunit has intrinsic tyrosine kinase activity that, upon insulin binding, activates intracellular signalling proteins by phosphorylating them on tyrosine residues¹¹⁵. These proteins include insulin-receptor substrate (IRS)-1, which is present in neurons and in peripheral tissues¹¹⁶. The neuronal protein encoded by the *Tub* gene is also tyrosine-phosphorylated in response to activation of the insulin receptor¹¹⁷. Because a loss-of-function mutation of *Tub* causes obesity in *tubby* mice, there is new support for the hypothesis that defective insulin signalling in the brain causes obesity.

Unlike insulin receptors, leptin receptors are members of the cytokine-receptor superfamily and do not have intrinsic tyrosine kinase activity¹¹⁸. However, the leptin receptor does have docking sites for janus kinases (JAK), a family of tyrosine kinases involved in intracellular cytokine signalling¹¹⁹. Activated JAK phosphorylates members of the signal transduction and transcription (STAT) family of intracellular proteins. STAT proteins, in turn, stimulate transcription of target genes that mediate some of leptin's cellular effects. Although

regulation of hypothalamic neuropeptide gene expression by leptin is well described, the role of JAK–STAT signalling in this response remains uncertain, as leptin can affect neuronal firing rate independently of its transcriptional effects. For example, a subset of 'glucose-responsive' neurons in the hypothalamus become hyperpolarized (and therefore decrease their firing rate) within minutes of leptin application¹²⁰. Glucose influences the membrane potential of glucose-responsive neurons indirectly through its oxidation to generate ATP, which in turn controls the activity of ATP-sensitive potassium channels (K_{ATP}) in the plasma membrane¹²¹. Closure of K_{ATP} channels in response to increasing intracellular concentrations of ATP (relative to ADP) raises intracellular $[K^+]$, which depolarizes the cell and increases its firing rate. Because leptin maintains K_{ATP} channels in the open configuration¹²⁰, positively charged K^+ ions diffuse out of the cell and lower its membrane potential. This effect on K_{ATP} channels can be detected even in isolated patches of plasma membrane, excluding a mechanism involving transcriptional effects of leptin. The relationships among leptin signalling through the JAK–STAT pathway, its effects on neuronal firing rate, and its control of neuropeptide gene expression are an important area for future study.



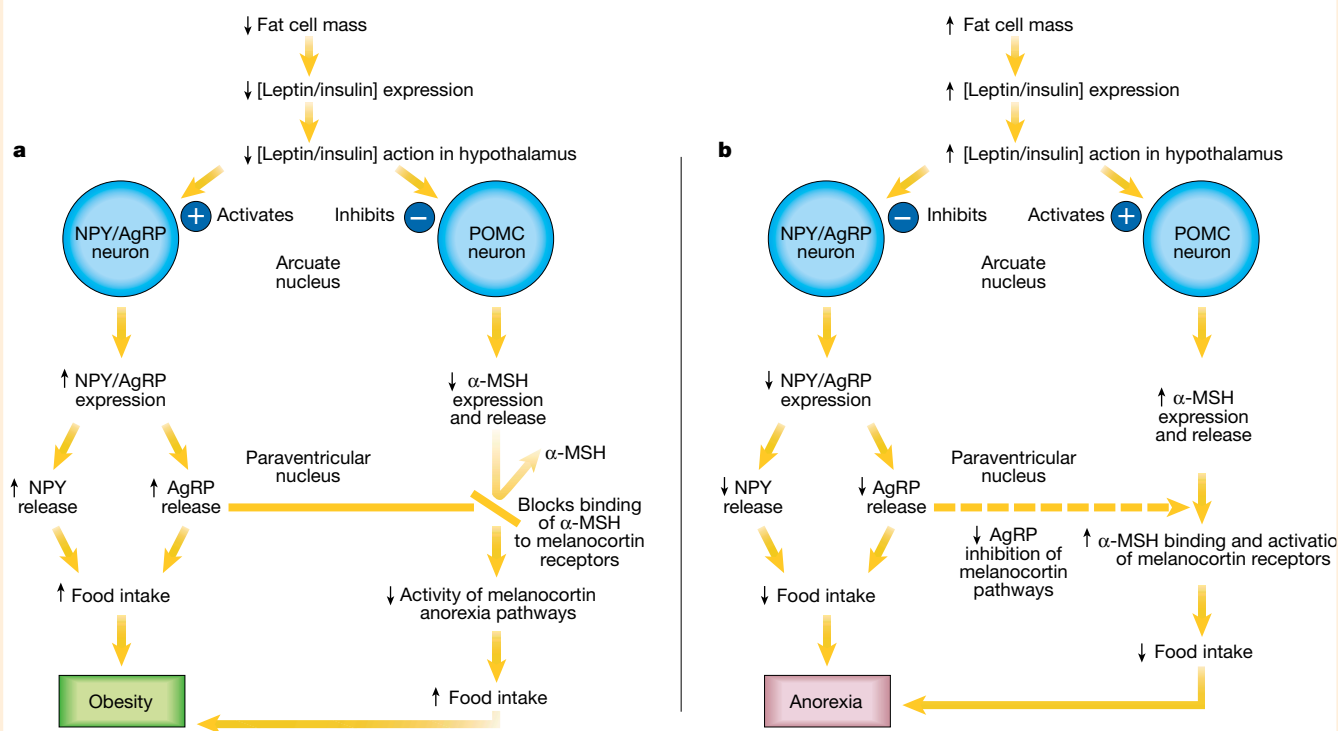


Figure 2 Role of arcuate nucleus neurons in adiposity signalling. **a**, Activity of leptin/insulin-sensitive adiposity signalling pathways in hypothalamus under conditions of leptin/insulin deficiency. **b**, Increased action of leptin/insulin in arcuate nucleus inhibits the NPY/AgRP anabolic pathway and stimulates the POMC catabolic pathway, resulting in reduced food intake and anorexia.

decreasing the ability of circulating leptin to enter brain interstitial fluid, where it can bind to neuronal leptin receptors, impaired leptin transport across endothelial cells of the blood–brain barrier is one potential mechanism. Several studies indicate that, like insulin⁹, leptin uptake into the brain is facilitated by leptin receptors expressed by endothelial cells³⁵ in the blood–brain barrier that function as leptin transporters. Whether dysfunction of this transport process can lead to obesity remains to be determined, but the finding that obese humans have leptin levels in cerebrospinal fluid that are low in comparison to plasma³⁶ is consistent with this possibility.

Reduced leptin-receptor signal transduction is another potential cause of leptin resistance. This has been documented not only in the brain of rodents bearing mutant leptin receptors, but also as an acquired response to leptin-receptor activation. Like some other cytokine receptors, activation of the leptin receptor induces expression of a protein that inhibits further leptin signal transduction, termed ‘suppressor of cytokine signalling-3’ (SOCS-3)³⁷. The potential contribution of SOCS-3 to acquired forms of leptin resistance and obesity is an active area of study.

Upon activation of leptin receptors in the brain, a series of integrated neuronal responses is probably required for food intake and energy balance to be affected. Failure of one or more neuronal systems in this circuit to respond to the leptin signal will therefore manifest as leptin resistance. The key role that these neuronal effector pathways have in energy homeostasis makes them an important priority for study and is the focus of the following discussion.

Neuropeptide effectors of adiposity signals

Several distinct hypothalamic neuropeptide-containing pathways have emerged as candidate mediators of leptin and insulin action in the CNS (Table 1).

Neuropeptide Y stimulates food intake

Prominent among anabolic effector pathways is a circuit containing

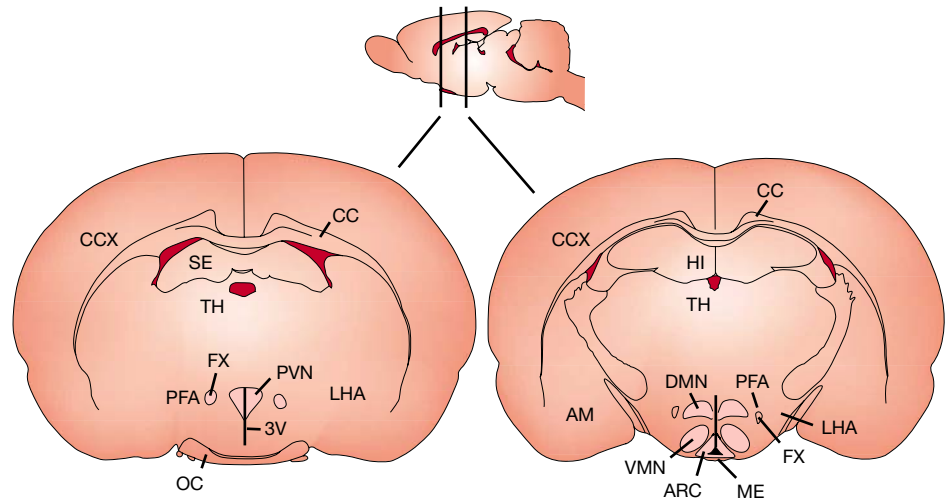
neuropeptide Y (NPY). Injection of NPY into cerebral ventricles or directly into the hypothalamus of rats potently stimulates food intake³⁸ and decreases energy expenditure while simultaneously inducing lipogenic enzymes in liver and white adipose tissue³⁹. Consequently, continuous or repeated central administration of NPY leads readily to obesity^{38,40}. Because NPY gene expression and secretion of the NPY peptide in the hypothalamus are increased during active depletion of body fat stores^{41,42} and/or reduced

Table 1 Neuropeptides implicated in the control of energy homeostasis

Molecule	Regulation by adiposity signals
Orexigenic	
NPY*	↓
AGRP*	↓
MCH	↓
Hypocretin 1 and 2/orexin A and B	↓
Galanin	?
Noradrenaline	?
Anorexigenic	
α-MSH†	↑
CRH*	↑
TRH*	↑
CART*	↑
IL-1β†	↑
Urocortin*	?
Glucagon-like peptide 1	?
Oxytocin	?
Neurotensin	?
Serotonin	?

Orexigenic refers to molecules that promote increased energy intake; anorexigenic implies the opposite. An asterisk designates documented, coordinated effects on both food intake and energy expenditure that promote a change in energy stores; arrows designate direction of effect exerted by one or both of the adiposity signals, leptin and insulin.

Figure 3 Diagrams of rat brain, showing major hypothalamic regions implicated in adiposity signalling and regulation of food intake. The small figure at the top is a longitudinal view of a rat brain, with olfactory bulb at the anterior end on the left and the caudal hindbrain on the right. Cross-sections of the brain at two levels (indicated by vertical lines) are shown at the left and right. First-order neurons responding to adiposity signals are located in the arcuate nucleus (ARC) and project anteriorly to the PVN as well as the PFA adjacent to the fornix (FX) and the LHA. Other regions implicated in regulating food intake include the ventromedial nucleus (VMN) and dorsomedial nucleus (DMN). Abbreviations of brain structures: AM, amygdala; CC, corpus callosum; CCX, cerebral cortex; HI, hippocampus; ME, median eminence; OC, optic chiasm; SE, septum; TH, thalamus; 3V, third ventricle.



leptin/insulin signalling to the brain⁴³, NPY meets the criteria for an anabolic signalling molecule. Moreover, leptin inhibits arcuate nucleus NPY gene expression^{44,45} and genetic knockout of NPY reduces hyperphagia and obesity in *ob/ob* mice⁴⁶, indicating that the full response to leptin deficiency requires NPY signalling (Fig. 2a). The hyperphagic response to insulin-deficient diabetes is similarly accompanied by increased hypothalamic NPY synthesis and release⁴⁷, and this response is blocked by insulin administration, either systemically or directly into the brain²⁰. The finding that mice that lack NPY (but are otherwise genetically normal) have intact feeding responses, however, raises questions about the need for NPY when leptin or insulin levels are normal⁴⁸. Alternatively, congenital absence of a major neuropeptide such as NPY may elicit compensatory responses that mask the consequences of its deficiency, and further study is required to resolve this issue. Agouti-related protein (AGRP), orexin (also known as 'hypocretin') and melanin-concentrating hormone (MCH) have subsequently been added to the list of candidate anabolic effector signalling molecules (Table 1).

Melanocortins suppress food intake

Candidate catabolic effector signalling molecules have an opposite set of characteristics. Melanocortins such as α -melanocyte-stimulating hormone (α -MSH)⁴⁹, as well as corticotropin-releasing hormone (CRH)⁵⁰, thyrotropin-releasing hormone (TRH)⁵¹, cocaine- and amphetamine-regulated transcript (CART)⁵² and interleukin-1 β ⁵³ are among a growing list of peptides that promote negative energy balance. Neuronal synthesis of these peptides increases in response to increased adiposity signalling in the brain. Among these, the melanocortin system stands out as being remarkable both for its complexity and its importance to energy homeostasis.

Melanocortins are peptides (such as α -MSH) that are cleaved from the pro-opiomelanocortin (POMC) precursor molecule and that exert their effects by binding to members of a family of melanocortin receptors⁴⁹. A role for melanocortin signalling in the control of energy homeostasis first emerged after the cloning of the MC3- and MC4-receptor genes and the demonstration that they are expressed primarily in the brain⁵⁴. This discovery was followed by evidence that a synthetic agonist of these receptors suppresses food intake, whereas a synthetic antagonist has the opposite effect⁵⁵. The report that mice lacking the MC4 receptor (owing to gene targeting) are hyperphagic and very obese⁵⁶ indicates that tonic signalling by MC4 receptors limits food intake and body fat mass. Mice heterozygous for the deleted MC4 allele also become obese, although less so than homozygous knockouts⁵⁶. Lack of a full complement of central

MC4 receptors, therefore, predisposes to hyperphagia and pathological weight gain. This finding has since been extended to humans with MC4-receptor mutations^{4,5}.

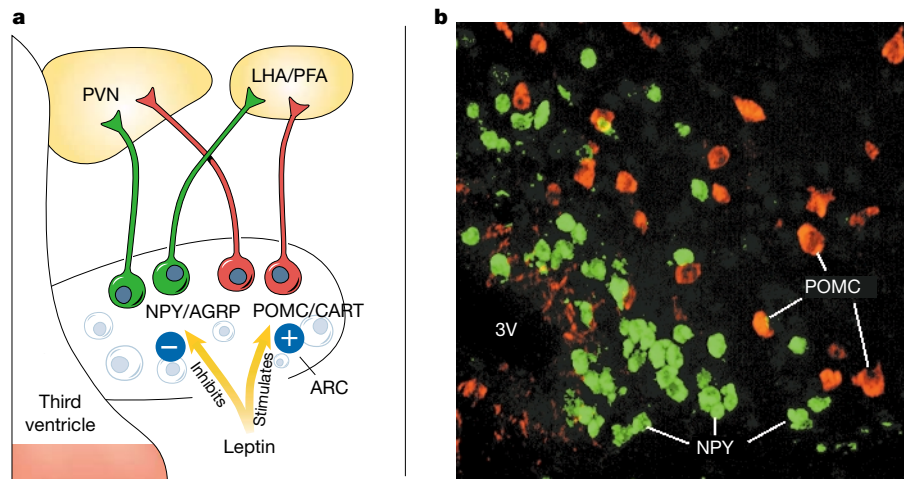
Further evidence for the importance of melanocortin signalling came from studies of agouti (*A^{y/a}*) mice, an autosomal dominant model of genetic obesity characterized by a yellow coat colour and an obese phenotype. Cloning of the *agouti* gene⁵⁷ identified a protein ('agouti') that functions as an antagonist of cutaneous MC1 receptors and normally is expressed by hair follicles. By reducing MC1 signalling, increased cutaneous agouti lightens the coat colour. Agouti mice, however, express agouti in tissues throughout the body and consequently develop both a yellow coat colour and obesity (owing to ectopic agouti production within the brain, where it antagonizes MC4 receptors)⁴⁹ (Fig. 2b).

The subsequent cloning of the *Agrp* gene⁵⁸ identified a peptide, AGRP, with homology to *agouti* that is an antagonist of MC3 and MC4 receptors⁵⁹. The demonstration that hypothalamic AGRP expression, like that of NPY and POMC, is localized to the arcuate nucleus⁵⁸, and that it is upregulated by fasting^{60,61} and by leptin deficiency⁵⁸, indicates that antagonism of CNS melanocortin receptors is important in body-weight regulation. Consistent with its role as an anabolic signalling molecule, AGRP causes hyperphagia when administered intracerebroventricularly (i.c.v.)^{62,63} or expressed transgenically⁵⁹, and the increase of food intake following a single i.c.v. injection of AGRP is sustained for up to a week⁶². Although NPY is described as the most potent orexigenic molecule (that is, a molecule that stimulates increased energy intake) when the feeding response is measured over a few hours, its effects are short-lived in comparison to those of AGRP. AGRP must therefore be considered the most robust orexigenic molecule if potency is measured as the cumulative increment of energy intake after a single i.c.v. injection. The mechanism underlying the extraordinary duration of action of AGRP remains a fascinating area for further investigation.

Neuropeptide signalling pathways in the hypothalamus

Brain lesioning and stimulation studies performed some six decades ago first implicated the hypothalamus as a major centre controlling food intake and body weight (Fig. 3). As summarized in a classic paper by Stellar⁶⁴, these studies identified the ventromedial hypothalamic nucleus (VMN) as the 'satiety centre', while the lateral hypothalamic area (LHA) was termed the 'hunger centre' (reviewed in ref. 50). These designations reflected the ability of electrical stimulation of the VMN to suppress food intake, and of bilateral VMN

Figure 4 NPY/AGRP and POMC/CART neurons in the arcuate nucleus, adjacent to the third ventricle, are first-order neurons in the hypothalamic response to the circulating adiposity signals insulin and leptin. **a**, Populations of first-order NPY/AGRP (green) and POMC/CART (red) neurons in the arcuate nucleus (ARC) are regulated by leptin and project to the PVN and to the LHA and PFA, which are locations of second-order hypothalamic neuropeptide neurons involved in the regulation of food intake and energy homeostasis. **b**, Fluorescence *in situ* hybridization detection of mRNAs encoding NPY (green cells) and POMC (red cells) in the arcuate nucleus, adjacent to the third ventricle (3V). NPY and POMC are expressed in discrete populations of arcuate nucleus neurons. NPY release in the PVN and LHA/PFA regions stimulates eating, whereas release of α -MSH (derived from POMC) in the PVN has an anorexic effect.



lesions to induce hyperphagia and obesity. Conversely, stimulation or lesioning of the LHA induced the opposite set of responses. As our knowledge of specific neuronal subpopulations involved in energy homeostasis has expanded, the notion of specific 'centres' of the brain that control food intake and body weight has been replaced by that of discrete neuronal pathways that generate integrated responses to afferent input related to changing body fuel stores⁶⁵.

Transduction of adiposity signals into a neuronal response

Situated adjacent to the floor of the third ventricle, the arcuate nucleus is an elongate ('arc-like') collection of neuronal cell bodies occupying approximately one-half of the length of the hypothalamus. NPY and AGRP are co-localized in arcuate nucleus neurons^{60,61}, demonstrating that a single neuronal cell type can contain multiple anabolic effector molecules. The subsequent finding that POMC and CART are co-localized in a distinct, but adjacent, subset of arcuate nucleus neurons⁶⁶ indicates that circuits originating in this brain area have highly specialized roles in energy homeostasis (Fig. 4).

The hypothesis that the arcuate nucleus transduces information related to signalling by leptin into a neuronal response is supported by the anorexic response to local microinjection of leptin into this area⁶⁷, and the inability of i.c.v. leptin to reduce food intake after the arcuate nucleus has been destroyed^{68,69}. A majority of both NPY/AGRP and POMC/CART neurons have been found to co-express leptin receptors^{16,17} and both types of neurons are regulated by leptin (as judged by changes in neuropeptide gene expression), but in an opposing manner. Thus, NPY/AGRP neurons are inhibited by leptin, and consequently are activated in conditions where leptin levels are low^{44,45,60,61}. Although less well studied, a deficiency of insulin also seems to activate these neurons^{20,47}, and insulin receptors are highly concentrated in the arcuate nucleus¹⁵. Conversely, conditions characterized by reduced insulin or leptin inhibit POMC^{70,71} and CART⁵² expression in the arcuate nucleus, and administration of these hormones can prevent or attenuate these neuropeptide responses. Moreover, involuntary overfeeding in rats, which potently inhibits spontaneous food intake once body weight has increased by more than 5%, elicits a threefold increase of POMC messenger RNA levels in the arcuate nucleus⁷². The demonstration that anorexia induced either by leptin⁷³ or by involuntary overfeeding⁷² is reversed by central administration of a melanocortin-receptor antagonist (at a low dose that has no effect on food intake in control animals) indicates that melanocortin signalling is a mediator of the anorexic response induced by increased adiposity signalling to the brain. Taken together, these findings indicate that the arcuate nucleus is a major site for transducing afferent input from circulating leptin and insulin into a neuronal response.

Implicit in this hypothesis is the suggestion that brain areas innervated by arcuate nucleus neurons are sites where second-order neurons involved in the energy homeostasis circuit are located. But the identification of such downstream neurons is just beginning, and energy homeostasis probably involves integrated and redundant pathways, rather than a discrete set of neurons connected in series to one another. Models for understanding how arcuate nucleus neurons ultimately affect food intake nonetheless provide a useful framework for future study. One possible model is proposed below.

Model for second-order neuronal signalling pathways

Hypothalamic areas including the paraventricular nucleus (PVN), zona incerta, perifornical area (PFA) and LHA are richly supplied by axons from arcuate nucleus NPY/AGRP and POMC/CART neurons^{74,75}. Insight into the role in energy homeostasis played by neurons in these areas can be gleaned from earlier stimulation and lesioning studies. For example, PVN stimulation inhibits food intake, whereas the reverse is true of stimulation of the LHA⁵⁰ and adjacent PFA⁷⁶. Conversely, bilateral PVN lesions cause a hyperphagic obesity syndrome, whereas bilateral lesioning of the LHA causes anorexia and weight loss^{50,64}. These observations indicate that anorexigenic and orexigenic signalling molecules might be synthesized in the PVN and LHA, respectively (Fig. 5).

Consistent with these predictions, several neuropeptides synthesized in PVN neurons reduce food intake and body weight when administered centrally. These include CRH, which causes anorexia and also activates the sympathetic nervous system in addition to its role as a major regulator of the hypothalamic-pituitary-adrenal axis^{50,77}; TRH, which reduces food intake⁵¹ in addition to stimulating the thyroid axis; and oxytocin, which reduces food intake in addition to regulating uterine function⁷⁸. If these PVN neurons are second-order catabolic effectors located downstream of the arcuate nucleus, they should be stimulated by melanocortin and/or CART signalling, but inhibited by NPY signalling, and further study is warranted to test this prediction.

The hypothesis that second-order neurons involved in anabolic signalling reside within the LHA/PFA is supported by studies of MCH, an orexigenic peptide located in this brain area⁷⁹. Evidence that MCH synthesis is elevated by both energy restriction and leptin deficiency⁷⁹, and that MCH-knockout mice have reduced food intake and are excessively lean⁸⁰, is consistent with this model. The discovery of the MCH receptor as a G-protein-coupled receptor (previously known as SLC-1)^{81,82} also supports the hypothesis of MCH as an orexigenic factor. Like NPY receptors, the MCH receptor is coupled to the G_i subunit of the plasma membrane G-protein assembly. By activating G_i, binding of MCH to its receptor inhibits formation of cyclic

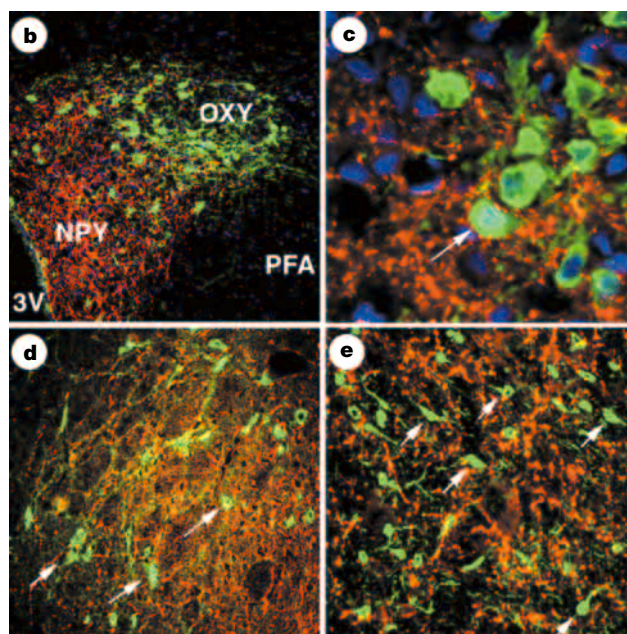
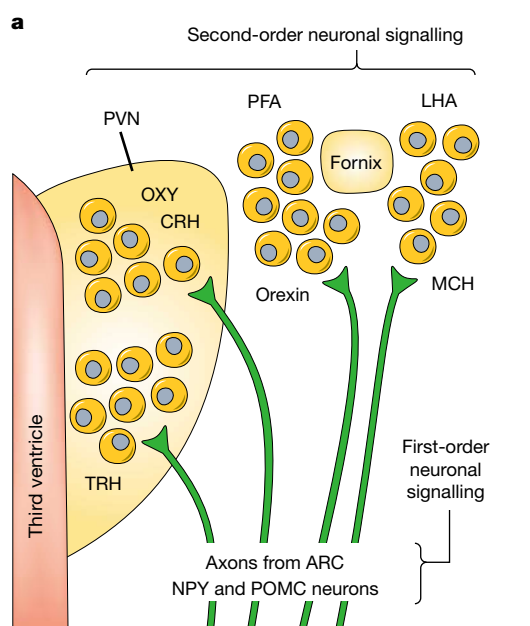


Figure 5 Locations of candidate second-order neurons involved in the hypothalamic response to insulin and leptin adiposity signalling. **a**, Diagram showing neuronal axons containing NPY and α -MSH from the arcuate nucleus (ARC) innervating the PVN, LHA and PFA (adjacent to the fornix). Candidate second-order neurons include those that express TRH, CRH and oxytocin (OXY) in the PVN (which cause anorexia), and neurons that express orexins and MCH in the PFA and LHA (which increase feeding). **b**, Fluorescence double-immunocytochemical staining of the PVN, showing NPY-containing axons (red fluorescence) in the parvocellular region adjacent to the third

ventricle and magnocellular OXY neurons (green fluorescence) in the lateral PVN. The PFA is located lateral to the PVN, but the fornix and LHA are not included in this field. **c**, Higher magnification of OXY neurons (green fluorescence) surrounded by axons and terminals containing NPY (red fluorescence). Nuclei of cells are shown in blue fluorescence. **d**, PFA showing NPY-containing axons (red fluorescence) surrounding neuron cell bodies containing orexins (green fluorescence). **e**, LHA showing NPY-containing axons (red fluorescence) surrounding neuron cell bodies containing MCH (green fluorescence).

AMP and consequently reduces signalling by protein kinase A (PKA)^{81,82}. This effect is opposite to that mediated by activation of receptors that exert anorexic effects, such as MC4 or CRH receptors, which are coupled to G_s and consequently increase cAMP and PKA signalling.

Two additional peptides are expressed exclusively in the LHA, zona incerta and PFA. Termed 'hypocretins 1 and 2'⁸³ or 'orexins A and B'⁸⁴ by the two groups that simultaneously discovered them, these peptides increase food intake and cause generalized behavioural arousal when administered centrally^{84,85}. Targeted deletion of the hypocretin/orexin gene in mice induces narcolepsy⁸⁶, a disorder characterized by the sudden onset of sleep at times when it would not ordinarily occur. This finding indicates that reduced hypocretin/orexin signalling may contribute to the onset and maintenance of sleep, in addition to its potential role in the control of food intake. Integration of MCH and hypocretin/orexin neurons into a model of the hypothalamic pathways controlling energy homeostasis predicts that they should be inhibited by melanocortin or CART input, and stimulated by NPY signalling, from neurons of the arcuate nucleus.

Much work must be done to test this model of first- and second-order neurons in the energy homeostasis circuit. Identifying specific neuronal subsets in the PVN and LHA that express NPY and melanocortin receptors is an important priority. Because many neurons of the PVN, PFA and LHA project to the arcuate nucleus, neuronal traffic flows bidirectionally between the arcuate nucleus and these other hypothalamic sites. So rather than being passive recipients of information from the arcuate nucleus, these second-order neurons can actively modify the information that arrives there. In addition, leptin receptors have been described on PVN and LHA neurons, implicating them as direct targets for regulation by circulating adiposity signals. However, far greater concentrations of leptin

receptors are present in the arcuate nucleus than in these other hypothalamic sites.

Satiety signals control meal size

It is self-evident that either the amount of food consumed during individual meals, the frequency of meals, or both, must be regulated if energy homeostasis is to be achieved. The major determinant of meal size is the onset of satiety, a biological state induced by neuro-humoral stimuli generated during food ingestion that leads to meal termination. To clarify how the decision to terminate a meal once it has begun is controlled in the regulation of energy homeostasis, we propose that hypothalamic pathways involved in energy homeostasis interact with a distinctly different set of pathways involved in the response to satiety signals^{65,87}. In contrast to the timing of meal initiation, which can be influenced by many external and internal variables (for example, emotional factors, time of day, availability and palatability of foods, and threats from the environment), meal termination tends to be a more biologically controlled process⁸⁸. Several findings indicate that control of meal size is a component of the feeding response induced by changes of body fuel stores or adiposity signalling. The hyperphagic response to central administration of NPY, for example, arises predominantly from the consumption of larger meals⁸⁹. Conversely, leptin-treated animals consume the same number of meals as vehicle-treated controls, but the meals are smaller⁹⁰. These observations indicate that signals involved in energy homeostasis may control food intake primarily by adjusting the size of individual meals. One way that this could be accomplished is by modulating the response to satiety signals in brain areas that process this information.

In contrast to its major role in mediating the response to adiposity signals, the hypothalamus is probably not the site that processes

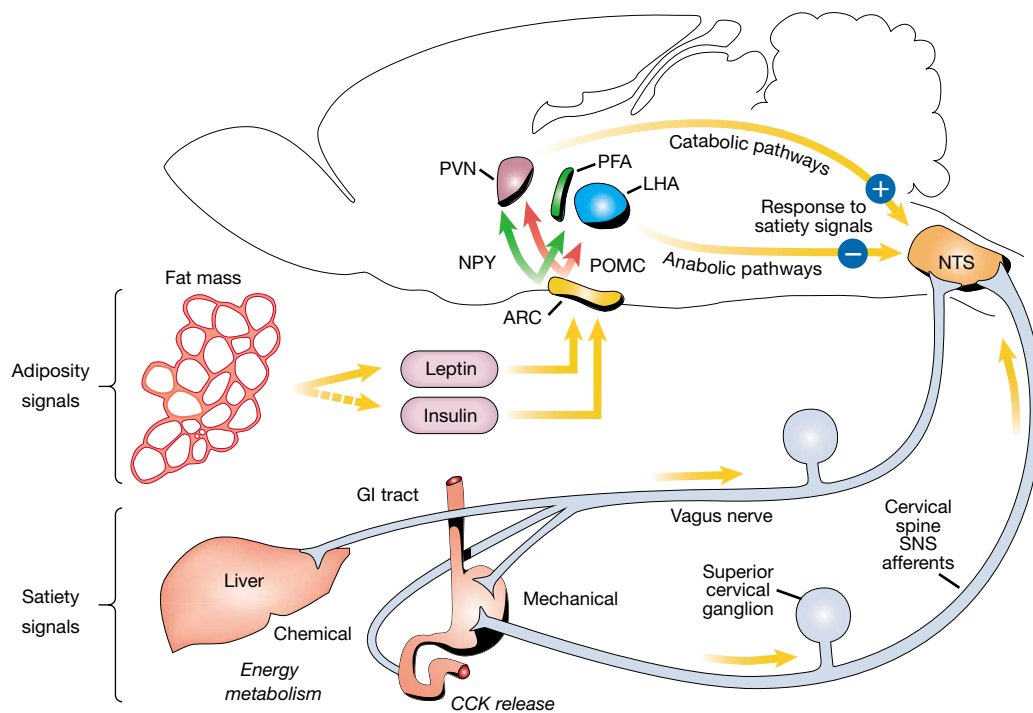


Figure 6 Neuroanatomical model of pathways by which adiposity signals, leptin (secreted by adipocytes) and insulin (secreted by the endocrine pancreas in proportion to adiposity), interact with central autonomic circuits regulating meal size. Leptin and insulin are proposed to stimulate a catabolic pathway (POMC/CART neurons) and inhibit an anabolic pathway (NPY/AGRP neurons) that originates in the arcuate nucleus (ARC). These pathways project to the PVN and LHA/PFA, where they make connections with central autonomic pathways that project to hindbrain autonomic centres that process satiety signals. Afferent input related to satiety from the liver, gastrointestinal tract and from peptides such as CCK are transmitted through the

vagus nerve and sympathetic fibres to the nucleus of the solitary tract (NTS), where they are integrated with descending hypothalamic input. Net neuronal output from the NTS and other hindbrain regions leads to the termination of individual meals, and is potentiated by catabolic projections from the PVN and inhibited by input from the LHA/PFA. Reduced input from adiposity signals (for example, during diet-induced weight loss), therefore, increases meal size by reducing brainstem responses to satiety signals. Not shown are ascending projections from hindbrain to forebrain that may also contribute to adaptive changes in food intake.

satiety signals. Rather, satiety information generated during the course of a meal is largely conveyed to the hindbrain by means of afferent fibres of the vagus nerve and by afferents passing into the spinal cord from the upper gastrointestinal tract⁹¹. This information converges in the nucleus tractus solitarius (NTS), an area of the caudal brainstem that integrates sensory information from the gastrointestinal tract and abdominal viscera, as well as taste information from the oral cavity⁹². Satiety-inducing signals that reach the NTS are initiated by mechanical or chemical stimulation of the stomach and small intestine during food ingestion, neural input related to energy metabolism in the liver⁹³ and humoral signals such as cholecystokinin (CCK) that are released upon nutrient stimulation of neuroendocrine secretory cells lining the intestinal lumen⁹⁴. Meal termination induced by such satiety signals can be demonstrated even when all neuronal connections between forebrain and hindbrain are severed⁹⁵. The basic process of terminating a meal, therefore, involves brain areas that can function in the absence of hypothalamic influences.

How then is the forebrain response to adiposity signals coupled to changes in the size of single meals? The hypothesis that such responses ultimately involve an interaction with hindbrain areas that control satiety is supported by the ability of both leptin⁹⁶ and insulin⁸⁷ to enhance the satiating effect of CCK. This interaction may be explained by the ability of central effector pathways to influence the response of NTS neurons to input from vagal afferents that convey satiety-related stimuli (Fig. 6). Recent evidence that leptin potentiates the effect of CCK to activate NTS

neurons demonstrates clearly that signals involved in energy homeostasis modulate the response of NTS neurons to input related to satiety⁹⁷.

Several caveats follow as a result from the hypothesis that NTS neurons are themselves responsible for integrating afferent information related to satiety with descending inputs from forebrain neurons involved in energy homeostasis. First, NTS neurons have reciprocal interconnections with forebrain areas such as the PVN⁹⁸, so the integration of satiety and energy homeostasis information probably involves multiple brain areas. In addition, the neuronal substrates for responding to central effector peptides involved in energy homeostasis are present locally within the NTS, as well as in the hypothalamus. For example, MC4 receptors are expressed in the NTS⁵⁴, and local administration of MC4-receptor agonists or antagonists into the fourth ventricle (which is adjacent to the NTS) elicits feeding responses that are indistinguishable from those induced by injecting these compounds into the more rostral lateral ventricle⁹⁹. This finding, combined with evidence that leptin receptors¹⁰⁰ and POMC neurons are both present in the NTS (the only brain area other than the arcuate nucleus that expresses the POMC gene)¹⁰¹, indicates that the hindbrain and forebrain may both process information involved in energy homeostasis. Thus, the NTS or other brainstem areas may, like the arcuate nucleus, contain neurons that respond to leptin and, through ascending projections to key forebrain sites, contribute to adaptive feeding responses to changes in body fat content. Further clarification of the mechanisms that integrate forebrain

and hindbrain circuitry involved in this process is an important area for future study.

Monoamine neurotransmitters and food intake

Noradrenaline

Noradrenaline is synthesized in brainstem areas such as the dorsal vagal complex and the locus ceruleus. These areas project both caudally to the spinal cord and rostrally to the hypothalamus, thalamus and cortex. In some of these neurons, including those projecting to the PVN, noradrenaline is co-localized with NPY. Like NPY, injection of noradrenaline into the PVN increases food intake robustly and repeated injections can result in substantial weight gain¹⁰². The observation of elevated noradrenaline levels in the PVN of *ob/ob* mice¹⁰³ indicates that leptin may inhibit noradrenaline release from terminals in this brain area, a possibility supported by *in vitro* studies using rat hypothalamus¹⁰⁴. Increased noradrenaline signalling in the PVN or other hypothalamic areas may therefore contribute to hyperphagia induced by leptin deficiency, a hypothesis that implicates noradrenaline as an anabolic effector in the CNS control of energy homeostasis.

Dopamine

A critical dependency of food intake on CNS dopamine signalling is implied by the profound feeding deficits that result from both pharmacological depletion¹⁰⁵ and genetic disruption¹⁰⁶ of dopamine synthesis. The interpretation of this finding is complicated, however, by motor impairments associated with dopamine deficiency that may also affect feeding behaviour. The observation that the feeding effects of dopamine vary with the brain region under study further obscures its role in energy homeostasis. For example, mesolimbic dopamine pathways (comprised of cell bodies in the substantia nigra and ventral tegmental area that project to the nucleus accumbens, striatum and cerebral cortex) seem to contribute to the 'rewarding' aspects of consuming palatable foods¹⁰⁷. In contrast, dopamine signalling in the hypothalamus via neurons situated in the dorsomedial and arcuate nuclei seems to inhibit food intake. Although reduced dopamine levels in the arcuate nucleus of *ob/ob* mice¹⁰³ raise the possibility that decreased hypothalamic dopamine signalling contributes to hyperphagia induced by leptin deficiency, the finding that leptin inhibits dopamine release from rat hypothalamus *in vitro*¹⁰⁴ is inconsistent with this hypothesis.

Synaptic concentrations of neurotransmitters are determined not only by the rate of their release from nerve terminals, but also by their rate of removal from the synaptic cleft. The latter process is dependent on specific transporter proteins that mediate neurotransmitter reuptake, the expression of which can be influenced by metabolic and hormonal factors¹⁰⁸. For example, fasting and uncontrolled diabetes reduce synaptic dopamine reuptake (which increases synaptic dopamine levels)¹⁰⁹, whereas the reverse is true for noradrenaline¹⁰⁸. Because these effects are reversed by insulin infusion directly into the brain, they may be mediated, at least in part, by reduced CNS insulin signalling.

Serotonin

The serotonin system is comprised of cell bodies in the caudal brainstem including the dorsal raphe nuclei that project widely throughout the brain, and is the primary target of several centrally acting drugs developed for obesity treatment (for example, dexfenfluramine and sibutramine). Such drugs increase serotonin-receptor signalling and thereby suppress food intake, whereas antagonists have the opposite effect¹¹⁰. The 5HT_{2C} serotonin-receptor subtype is implicated in this process, as knockout of this receptor increases food intake and body weight¹¹¹. Maintenance of normal energy homeostasis, therefore, seems to require intact serotonin signalling. But obesity in this model is modest, especially when compared to the phenotype of mice lacking MC4 or leptin receptors. The recent finding that leptin increases serotonin turnover¹¹² raises the possibility that at least some of leptin's weight-reducing effects are mediated through increased serotonin signalling. However, leptin-induced anorexia is intact in mice lacking

the 5HT_{2C} receptor¹¹¹, indicating that leptin's ability to reduce food intake does not require signalling at this receptor subtype.

Aminergic neurotransmitter systems, therefore, exert unambiguous effects on food intake and provide important targets for current approaches to drug treatment of obesity. The role of these systems in energy homeostasis is complex, however, and the hypothesis that they serve as major targets for the action of adiposity signals is not strongly supported.

Therapeutic implications

A more detailed understanding of the pathogenesis of human obesity may ultimately guide treatment of affected individuals. Obesity that results from reduced melanocortins, for example, might respond well to administration of melanocortin-receptor agonists, if and when they become available in a clinical setting. Evidence for this is provided in a recent study of POMC-knockout mice, in which obesity was reversed by administration of an MC4-receptor agonist¹¹³. Analogously, administering leptin to an obese human with genetic leptin deficiency reduced weight as markedly as it does in *ob/ob* mice¹¹⁴. Obesity that is associated with leptin resistance, however, may be common and would be unlikely to respond to leptin treatment unless the resistance can be overcome. Patients with defective melanocortin-receptor function, for example, seem unlikely to respond to therapy with either leptin or melanocortin-receptor agonists. These considerations indicate that an expanded ability to diagnose the pathophysiological basis of human obesity will have direct applications to its treatment. However, a multi-drug regimen that targets multiple sites within the weight-regulatory system may be necessary to achieve and sustain weight loss in many individuals.

The impressive effects of AGRP on food intake in rodents^{62,63} indicate that it warrants evaluation in the treatment of conditions associated with excessive weight loss, including anorexia nervosa and wasting illness associated with AIDS or cancer. If AGRP proves ineffective in this context, it would indicate that the pathogenesis of such conditions involves pathways additional to, and potentially downstream of, melanocortin-receptor activation. Such an outcome might therefore direct therapeutic strategies towards activation of other candidate anabolic (or inhibition of catabolic) effector pathways.

These considerations highlight the importance of clarifying the mechanisms that control food intake and energy homeostasis. Such information will help us to understand the pathogenesis of disorders at both ends of the body-weight spectrum, and is a probable prerequisite for their successful treatment. The enormous cost to human health attributable to these disorders emphasizes the need for a more complete understanding of this area. □

1. Edholm, O. G. Energy balance in man. *J. Human Nutr.* **31**, 413–431 (1977).
2. Montague, C. *et al.* Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* **387**, 903–908 (1997).
3. Clement, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 330–331 (1998).
4. Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nature Genet.* **20**, 113–114 (1998).
5. Yeo, G. *et al.* A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nature Genet.* **20**, 111–112 (1998).
6. Harris, R. B. S., Kasser, T. R. & Martin, R. J. Dynamics of recovery of body composition after overfeeding, food restriction or starvation of mature female rats. *J. Nutr.* **116**, 2536–2546 (1986).
7. Kennedy, G. C. The role of depot fat in the hypothalamic control of food intake in the rat. *Proc. R. Soc. Lond. B* **140**, 579–592 (1953).
8. Gibbs, J., Young, R. C. & Smith, G. P. Cholecystokinin decreases food intake in rats. *J. Comp. Physiol. Psychol.* **84**, 488–495 (1973).
9. Baura, G. *et al.* Saturable transport of insulin from plasma into the central nervous system of dogs in vivo: a mechanism for regulated insulin delivery to the brain. *J. Clin. Invest.* **92**, 1824–1830 (1993).
10. Woods, S., Lotter, E., McKay, L. & Porte, D. J. Chronic intracerebroventricular infusion of insulin reduces food intake and body weight of baboons. *Nature* **282**, 503–505 (1979).
11. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
12. Bagdade, J. D., Bierman, E. L. & Porte, D. Jr The significance of basal insulin levels in the evaluation of the insulin response to glucose in diabetic and nondiabetic subjects. *J. Clin. Invest.* **46**, 1549–1557 (1967).
13. Considine, R. V. *et al.* Serum immunoreactive-leptin concentrations in normal-weight and obese humans. *New Eng. J. Med.* **334**, 292–295 (1996).
14. Schwartz, M. W. *et al.* Cerebrospinal fluid leptin levels: relationship to plasma levels and to adiposity

- in humans. *Nature Med.* **2**, 589–593 (1996).
15. Baskin, D. G., Wilcox, B. J., Figlewicz, D. P. & Dorsa, D. M. Insulin and insulin-like growth factors in the CNS. *Trends Neurosci.* **11**, 107–111 (1988).
16. Baskin, D., Breininger, J. & Schwartz, M. Leptin receptor mRNA identifies a subpopulation of neuropeptide Y neurons activated by fasting in rat hypothalamus. *Diabetes* **48**, 828–833 (1999).
17. Cheung, C., Clifton, D. & Steiner, R. Proopiomelanocortin neurons are direct targets for leptin in the hypothalamus. *Endocrinology* **138**, 4489–4492 (1997).
18. Weigle, D. S. *et al.* Recombinant *ob* protein reduces feeding and body weight in the *ob/ob* mouse. *J. Clin. Invest.* **96**, 2065–2070 (1995).
19. Campfield, L., Smith, F., Gulsez, Y., Devos, R. & Burn, P. Mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks. *Science* **269**, 546–549 (1995).
20. Sipols, A. J., Baskin, D. G. & Schwartz, M. W. Effect of intracerebroventricular insulin infusion on diabetic hyperphagia and hypothalamic neuropeptide gene expression. *Diabetes* **44**, 147–151 (1995).
21. Porte, D. J., Woods, S., Baskin, D., Seeley, R. & Schwartz, M. Obesity, diabetes and the central nervous system. *Diabetologia* **41**, 863–881 (1998).
22. Schwartz, M. W. *et al.* Evidence that plasma leptin and insulin levels are associated with body adiposity via different mechanisms. *Diabetes Care* **20**, 1476–1481 (1997).
23. Kahn, S. E. *et al.* Quantification of the relationship between insulin sensitivity and B-cell function in human subjects: evidence for a hyperbolic function. *Diabetes* **42**, 1663–1672 (1993).
24. Polonsky, K. S., Given, B. D. & VanCauter, E. Twenty-four hour profiles and pulsatile patterns of insulin secretion in normal and obese subjects. *J. Clin. Invest.* **81**, 442–448 (1988).
25. Mueller, W. *et al.* Evidence that glucose metabolism regulates leptin secretion from cultured rat adipocytes. *Endocrinology* **139**, 551–558 (1998).
26. Wang, J., Liu, R., Hawkins, M., Barzilai, N. & Rossetti, L. A nutrient-sensing pathway regulates leptin gene expression in muscle and fat. *Nature* **393**, 684–688 (1998).
27. Leedom, L. J. & Meehan, W. P. The psychoneuroendocrinology of diabetes mellitus in rodents. *Psychoneuroendocrinology* **14**, 275–294 (1989).
28. Havel, P. *et al.* Marked and rapid decreases of circulating leptin in streptozotocin diabetic rats: reversal by insulin. *Am. J. Physiol.* **274**, R1482–R1491 (1998).
29. Hathout, E. *et al.* Changes in plasma leptin during the treatment of diabetic ketoacidosis. *J. Clin. Endocrinol. Metab.* **84**, 4545–4548 (1999).
30. Sindelar, D. *et al.* Low plasma leptin levels contribute to diabetic hyperphagia in rats. *Diabetes* **48**, 1275–1280 (1999).
31. Chua, S. C. *et al.* Phenotypes of mouse diabetes and rat fatty due to mutations in the OB (Leptin) receptor. *Science* **271**, 994–996 (1996).
32. Mantzoros, C. *et al.* Severe leptin resistance in brown fat-deficient uncoupling protein promoter-driven diphtheria toxin A mice despite suppression of hypothalamic neuropeptide Y and circulating corticosterone. *Diabetes* **47**, 230–238 (1998).
33. Marsh, D. J. *et al.* Response of melanocortin-4 receptor-deficient mice to anorectic and orexigenic peptides. *Nature Genet.* **21**, 119–122 (1999).
34. Wilson, B. *et al.* Physiological and anatomical circuitry between Agouti-related protein and leptin signaling. *Endocrinology* **140**, 2387–2397 (1999).
35. Bjorbaek, C. *et al.* Expression of leptin receptor isoforms in rat brain microvessels. *Endocrinology* **139**, 3485–3491 (1998).
36. Caro, J. F. *et al.* Decreased cerebrospinal-fluid/serum leptin ratio in obesity: a possible mechanism for leptin resistance. *Lancet* **348**, 159–161 (1996).
37. Bjorbaek, C., Elmquist, J., Frantz, J., Shoelson, S. & Flier, J. Identification of SOCS-3 as a potential mediator of central leptin resistance. *Mol. Cell* **1**, 619–625 (1998).
38. Stanley, B. G., Kyrkouli, S. E., Lampert, S. & Leibowitz, S. F. Neuropeptide Y chronically injected into the hypothalamus: a powerful neurochemical inducer of hyperphagia and obesity. *Peptides* **7**, 1189–1192 (1986).
39. Billington, C. J., Briggs, J. E., Grace, M. & Levine, A. S. Effects of intracerebroventricular injection of neuropeptide Y on energy metabolism. *Am. J. Physiol.* **260**, R321–R327 (1991).
40. Zazrevski, N., Cusin, I., Vetter, R., Rohner-Jeanraud, F. & Jeanraud, B. Chronic intracerebroventricular neuropeptide-Y administration to normal rats mimics hormonal and metabolic changes of obesity. *Endocrinology* **133**, 1753–1758 (1993).
41. Kalra, S. P., Dube, M. G., Sahu, A., Phelps, C. P. & Kalra, P. Neuropeptide Y secretion increases in the paraventricular nucleus in association with increased appetite for food. *Proc. Natl Acad. Sci. USA* **88**, 10931–10935 (1991).
42. White, J. D. & Kershaw, M. Increased hypothalamic neuropeptide Y expression following food deprivation. *Mol. Cell. Neurosci.* **1**, 41–48 (1989).
43. Wilding, J. P. H. *et al.* Increased neuropeptide-Y messenger ribonucleic acid (mRNA) and decreased neurotensin mRNA in the hypothalamus of the obese (*ob/ob*) mouse. *Endocrinology* **132**, 1939–1944 (1993).
44. Schwartz, M. W. *et al.* Specificity of leptin action on elevated blood glucose levels and hypothalamic neuropeptide Y gene expression in *ob/ob* mice. *Diabetes* **45**, 531–535 (1996).
45. Stephens, T. W. *et al.* The role of neuropeptide Y in the antiobesity action of the obese gene product. *Nature* **377**, 530–534 (1995).
46. Erickson, J. C., Holloper, G. & Palmiter, R. D. Attenuation of the obesity syndrome of *ob/ob* mice by the loss of neuropeptide Y. *Science* **274**, 1704–1707 (1996).
47. Williams, G. *et al.* Increased neuropeptide Y concentrations in specific hypothalamic regions of streptozotocin-induced diabetic rats. *Diabetes* **38**, 321–327 (1989).
48. Erickson, J. C., Clegg, K. E. & Palmiter, R. D. Sensitivity to leptin and susceptibility to seizures of mice lacking neuropeptide Y. *Nature* **381**, 415–418 (1996).
49. Cone, R. *et al.* The melanocortin receptors: agonists, antagonists, and the hormonal control of pigmentation. *Rec. Prog. Horm. Res.* **51**, 287–318 (1996).
50. Bray, G. A., Fisler, J. & York, D. A. Neuroendocrine control of the development of obesity: understanding gained from studies of experimental animal models. *Front. Neuroendocrinol.* **11**, 128–181 (1990).
51. Kow, L. & Pfaff, D. The effects of the TRH-metabolite cyclo(His-Pro) and its analogs on feeding. *Pharmacol. Biochem. Behav.* **38**, 359–364 (1991).
52. Kristensen, P. *et al.* Hypothalamic CART is a new anorectic peptide regulated by leptin. *Nature* **393**, 72–76 (1998).
53. Lusheski, G., Gardner, J., Rushforth, D., Loudon, A. & Rothwell, N. Leptin actions on food intake and body temperature are mediated by IL-1. *Proc. Natl Acad. Sci. USA* **96**, 7047–7052 (1999).
54. Mountjoy, K., Mortrud, M., Low, M., Simerly, R. & Cone, R. Localization of the melanocortin-4 receptor (MC4R) in neuroendocrine and autonomic control circuits in the brain. *Mol. Endocrinol.* **8**, 1298–1308 (1994).
55. Fan, W., Boston, B., Kesterson, R., Hruby, V. & Cone, R. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* **385**, 165–168 (1997).
56. Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131–141 (1997).
57. Miller, M. *et al.* Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutation. *Genes Dev.* **7**, 454–467 (1993).
58. Shutter, J. *et al.* Hypothalamic expression of ART, a novel gene related to agouti, is up-regulated in obese and diabetic mutant mice. *Genes Dev.* **11**, 593–602 (1997).
59. Ollmann, M. *et al.* Antagonism of central melanocortin receptors in vitro and in vivo by agouti-related protein. *Science* **278**, 135–138 (1997).
60. Hahn, T., Breininger, J., Baskin, D. & Schwartz, M. Coexpression of AgRP and NPY in fasting-activated hypothalamic neurons. *Nature Neurosci.* **1**, 271–272 (1998).
61. Broberger, C., Johansen, J., Johansson, C., Schalling, M. & Hokfelt, T. The neuropeptide Y/agouti gene-related protein (AGRP) brain circuitry in normal, anorectic, and monosodium glutamate-treated mice. *Proc. Natl Acad. Sci. USA* **95**, 15043–15048 (1998).
62. Hagan, M. *et al.* Effect of agouti-related peptide (83–132) on food intake: competitive melanocortin-receptor antagonism on short but not long term effects. *Am. J. Physiol.* (in the press).
63. Rossi, M. *et al.* A C-terminal fragment of Agouti-related protein increases feeding and antagonizes the effect of alpha-melanocyte stimulating hormone in vivo. *Endocrinology* **139**, 4428–4431 (1998).
64. Stellar, E. The physiology of motivation. *Psychol. Rev.* **61**, 5 (1954).
65. Woods, S., Seeley, R., Porte, D. J. & Schwartz, M. Signals that regulate food intake and energy homeostasis. *Science* **280**, 1378–1383 (1998).
66. Elias, C. *et al.* Leptin activates hypothalamic CART neurons projecting to the spinal cord. *Neuron* **21**, 1375–1385 (1998).
67. Satoh, N. *et al.* The arcuate nucleus as a primary site of satiety effect of leptin in rats. *Neurosci. Lett.* **224**, 149–152 (1997).
68. Tang-Christensen, M., Holst, J., Hartmann, B. & Vrang, N. The arcuate nucleus is pivotal in mediating the anorectic effects of centrally administered leptin. *Neuroreport* **10**, 1183–1187 (1999).
69. Dawson, R., Pellemounter, M., Millard, W., Liu, S. & Eppler, B. Attenuation of leptin-mediated effects by monosodium glutamate-induced arcuate nucleus damage. *Am. J. Physiol.* **273**, E202–E206 (1997).
70. Schwartz, M. *et al.* Leptin increases hypothalamic proopiomelanocortin (POMC) mRNA expression in the rostral arcuate nucleus. *Diabetes* **46**, 2119–2123 (1997).
71. Thornton, J., Cheung, C., Clifton, D. & Steiner, R. Regulation of hypothalamic proopiomelanocortin mRNA by leptin in *ob/ob* mice. *Endocrinology* **138**, 5063–5066 (1997).
72. Hagan, M. *et al.* Role of the CNS melanocortin system in the response to overfeeding. *J. Neurosci.* **19**, 2362–2367 (1999).
73. Seeley, R. *et al.* Melanocortin receptors in leptin effects. *Nature* **390**, 349 (1997).
74. Elmquist, J., Maratos-Flier, E., Saper, C. & Flier, J. Unraveling the central nervous system pathways underlying responses to leptin. *Nature Neurosci.* **1**, 445–450 (1998).
75. Elmquist, J., Elias, C. & Saper, C. From lesions to leptin: hypothalamic control of food intake and body weight. *Neuron* **22**, 221–232 (1999).
76. Stanley, B. G., Willett, V. L. III, Donias, H. W., Ha, L. H. & Spears, L. C. The lateral hypothalamus: a primary site mediating excitatory amino acid-elicited eating. *Brain Res.* **630**, 41–49 (1993).
77. Dallman, M. F. *et al.* Feast and famine: Critical role of glucocorticoids with insulin in daily energy flow. *Front. Neuroendocrinol.* **14**, 303–347 (1993).
78. Verbalis, J., Blackburn, R., Hoffman, G. & Stricker, E. Establishing behavioral and physiological functions of central oxytocin: insights from studies of oxytocin and ingestive behaviors. *Adv. Exp. Med. Biol.* **395**, 209–225 (1995).
79. Qu, D. *et al.* A role for melanin-concentrating hormone in the central regulation of feeding behaviour. *Nature* **380**, 243–247 (1996).
80. Shimada, M., Tritos, N., Lowell, B., Flier, J. & Maratos-Flier, E. Mice lacking melanin-concentrating hormone are hypophagic and lean. *Nature* **39**, 670–674 (1998).
81. Saito, Y. *et al.* Molecular characterization of the melanin-concentrating-hormone receptor. *Nature* **400**, 265–269 (1999).
82. Chambers, J. *et al.* Melanin-concentrating hormone is the cognate ligand for the orphan G-protein-coupled receptor SLC-1. *Nature* **400**, 261–265 (1999).
83. de Lecea, L. *et al.* The hypocretins: hypothalamus-specific peptides with neuroexcitatory activity. *Proc. Natl Acad. Sci. USA* **95**, 322–327 (1998).
84. Sakurai, T. *et al.* Orexins and orexin receptors: a family of hypothalamic neuropeptides and G-protein-coupled receptors that regulate feeding behavior. *Cell* **92**, 573–585 (1998).
85. Hagan, J. *et al.* Orexin A activates locus coeruleus cell firing and increases arousal in the rat. *Proc. Natl Acad. Sci. USA* **96**, 10911–10916 (1999).
86. Chemelli, R. *et al.* Narcolepsy in orexin knockout mice: molecular genetics of sleep regulation. *Cell* **98**, 437–451 (1999).
87. Figlewicz, D. P., Stein, L. J., West, D., Porte, D. Jr & Woods, S. C. Intracisternal insulin alters sensitivity to CCK-induced meal suppression in baboons. *Am. J. Physiol.* **250**, R856–R860 (1986).
88. Woods, S. C. & Strubbe, J. H. The psychobiology of meals. *Psychonomic Bull. Rev.* **1**, 141–155 (1994).
89. Leibowitz, S. & Alexander, J. Analysis of neuropeptide Y-induced feeding: dissociation of Y1 and Y2 receptor effects on natural meal patterns. *Peptides* **12**, 1251–1260 (1991).
90. Flynn, M., Scott, T., Pritchard, T. & Plata-Salamán, C. Mode of action of OB protein (leptin) on feeding. *Am. J. Physiol.* **275**, R174–R179 (1998).
91. Ritter, S., Dinh, T. & Friedman, M. Induction of Fos-like immunoreactivity (Fos-li) and stimulation of feeding by 2,5-anhydro-D-mannitol (2,5-AM) require the vagus nerve. *Brain Res.* **646**, 53–64 (1994).
92. Travers, S. & Norgren, R. Gustatory neural processing in the hindbrain. *Annu. Rev. Neurosci.* **10**, 595–632 (1987).
93. Friedman, M., Harris, R., Ji, H., Ramirez, I. & Tordoff, M. Fatty acid oxidation affects food intake by altering hepatic energy status. *Am. J. Physiol.* **276**, R1046–R1053 (1999).
94. Moran, T. & Schwartz, G. Neurobiology of cholecystokinin. *Crit. Rev. Neurobiol.* **9**, 1–28 (1994).
95. Grill, H. & Smith, G. Cholecystokinin decreases sucrose intake in chronic decerebrate rats. *Am. J. Physiol.* **254**, R853–R854 (1988).

96. Matson, C. & Ritter, R. Long-term CCK-leptin synergy suggests a role for CCK in the regulation of body weight. *Am. J. Physiol.* **276**, R1038–R1045 (1999).
97. Emond, M., Schwartz, G., Ladenheim, E. & Moran, T. Central leptin modulates behavioral and neural responsivity to CCK. *Am. J. Physiol.* **276**, R1545–R1549 (1999).
98. Ter Horst, G., de Boer, P., Luiten, P. & van Willigen, J. Ascending projections from the solitary tract nucleus to the hypothalamus. A *Phaseolus vulgaris* lectin tracing study in the rat. *Neuroscience* **31**, 785–797 (1989).
99. Grill, H., Ginsberg, A., Seeley, R. & Kaplan, J. Brainstem application of melanocortin receptor ligands produces long-lasting effects on feeding and body weight. *J. Neurosci.* **18**, 10128–10135 (1998).
100. Mercer, J., Moar, K. & Hoggard, N. Localization of leptin receptor (Ob-R) messenger ribonucleic acid in the rodent hindbrain. *Endocrinology* **139**, 29–34 (1998).
101. Bronstein, D., Schafer, M., Watson, S. & Akil, H. Evidence that beta-endorphin is synthesized in cells in the nucleus tractus solitarius: detection of POMC mRNA. *Brain Res.* **587**, 269–275 (1992).
102. Leibowitz, S., Roosin, P. & Rosenn, M. Chronic norepinephrine injection into the hypothalamic paraventricular nucleus produces hyperphagia and increased body weight in the rat. *Pharmacol. Biochem. Behav.* **21**, 801–808 (1984).
103. Olthmans, G. Norepinephrine and dopamine levels in hypothalamic nuclei of the genetically obese mouse (*ob/ob*). *Brain Res.* **273**, 369–373 (1983).
104. Brunetti, L., Michelotto, B., Orlando, G. & Vacca, M. Leptin inhibits norepinephrine and dopamine release from rat hypothalamic neuronal endings. *Eur. J. Pharmacol.* **372**, 237–240 (1999).
105. Salamone, J., Mahan, K. & Rogers, S. Ventrolateral striatal dopamine depletions impair feeding and food handling in rats. *Pharmacol. Biochem. Behav.* **44**, 605–610 (1993).
106. Szczypka, M. *et al.* Feeding behavior in dopamine-deficient mice. *Proc. Natl Acad. Sci. USA* **96**, 12138–12143 (1999).
107. Pothos, E., Creese, I. & Hoebel, B. Restricted eating with weight loss selectively decreases extracellular dopamine in the nucleus accumbens and alters dopamine response to amphetamine, morphine, and food intake. *J. Neurosci.* **15**, 6640–6650 (1995).
108. Figlewicz, D., Endocrine regulation of neurotransmitter transporters. *Epilepsy Res.* **37**, 203–210 (1999).
109. Patterson, T. *et al.* Food deprivation decreases mRNA and activity of the rat dopamine transporter. *Neuroendocrinology* **68**, 11–20 (1998).
110. Leibowitz, S. & Alexander, J. Hypothalamic serotonin in control of eating behavior, meal size, and body weight. *Biol. Psychiatry* **44**, 851–864 (1998).
111. Nonogaki, K., Strack, A., Dallman, M. & Tecott, L. Leptin-independent hyperphagia and type 2 diabetes in mice with a mutated serotonin 5-HT_{2C} receptor gene. *Nature Med.* **4**, 1152–1156 (1998).
112. Calapai, G. *et al.* Leptin increases serotonin turnover by inhibition of brain nitric oxide synthesis. *J. Clin. Invest.* **104**, 975–982 (1999).
113. Yaswen, L., Diehl, N., Brennam, M. & Hochgeschwender, V. Obesity in the mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin. *Nature Med.* **5**, 1066–1070 (1999).
114. Farooqi, I. *et al.* Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *N. Engl. J. Med.* **341**, 879–884 (1999).
115. Cheatham, B. & Kahn, C. Insulin action and the insulin signaling network. *Endocr. Rev.* **16**, 117–142 (1995).
116. Baskin, D. G., Sipols, A. J., Schwartz, M. W. & White, M. F. Insulin receptor substrate-1 (IRS-1) expression in rat brain. *Endocrinology* **134**, 1952–1955 (1994).
117. Kapeller, R. *et al.* Tyrosine phosphorylation of tub and its association with Src homology 2 domain-containing proteins implicate tub in intracellular signaling by insulin. *J. Biol. Chem.* **274**, 24980–24986 (1999).
118. Tartaglia, L. A. *et al.* Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**, 1263–1271 (1995).
119. Tartaglia, L. A. The leptin receptor. *J. Biol. Chem.* **272**, 6093–6096 (1997).
120. Spanswick, D., Smith, M., Groppi, V., Logam, S. & Ashford, M. Leptin inhibits hypothalamic neurons by activation of ATP-sensitive potassium channels. *Nature* **390**, 521–525 (1997).
121. Dunn-Meynell, A., Rawson, N. & Levin, B. Distribution and phenotype of neurons containing the ATP-sensitive K⁺ channel in rat brain. *Brain Res.* **814**, 41–54 (1998).

Acknowledgements

Assistance in manuscript preparation provided by L. Walters and in figure preparation by M. Baskin is gratefully acknowledged. This work was supported by NIH grants, the Diabetes Endocrinology Research Center and Clinical Nutrition Research Unit of the University of Washington, and the Department of Veterans Affairs.



Medicinal strategies in the treatment of obesity

George A. Bray* & Louis A. Tartaglia†

*Pennington Biomedical Research Center, Baton Rouge, Louisiana 70808, USA (e-mail: BrayGA@pbrc.edu)

†Millennium Pharmaceuticals Inc., 75 Sidney Street, Cambridge, Massachusetts 02139, USA

When prevention fails, medicinal treatment of obesity may become a necessity. Any strategic medicinal development must recognize that obesity is a chronic, stigmatized and costly disease that is increasing in prevalence. Because obesity can rarely be cured, treatment strategies are effective only as long as they are used, and combined therapy may be more effective than monotherapy. For a drug to have significant impact on body weight it must ultimately reduce energy intake, increase energy expenditure, or both. Currently approved drugs for long-term treatment of obesity include sibutramine, which inhibits food intake, and orlistat, which blocks fat digestion.

Breakthroughs in our understanding of the molecular mechanisms regulating body weight have provided potential opportunities for therapeutic intervention and brought renewed hope and vitality to the development of antiobesity drugs. Several therapeutic agents are around the corner and still others are on the horizon. In this review we will discuss potential strategies for developing new drugs and specific antiobesity targets representing a number of diverse mechanisms.

Obesity is a chronic, stigmatized and costly disease that is rarely curable and is increasing in prevalence in most of the world^{1,2}. It is estimated to cause between 280,000 and 325,000 deaths per year in the United States³. At this time, available treatments, including drugs, are palliative and are effective only while the treatment is being actively used⁴; when effective drugs or other treatments are discontinued, weight gain is an inevitable consequence. Any effective drug will be widely used for its cosmetic as well as medical benefits, as obesity is a socially stigmatized problem. Thus, any drug that is approved should meet high standards for safety. Because treating obesity results in only a 10% weight loss, many patients are dissatisfied and doctors would prefer to treat the hypertension, diabetes or dyslipidaemia that is

normally associated with obesity. Delays in treatment increase the risk of future development of diabetes and its related complications, and of heart disease. The latter effect would reverse the recent favourable downward trends with this disease.

Mechanisms for drug action

Many features of obesity are analogous to hypertension^{4,5}. Both diseases can be described by feedback systems in which knowledge about the underlying mechanisms has increased markedly in the past 20–30 years.

Multiple sites in this feedback system have proven useful for developing antihypertensive drugs, and we anticipate that multiple sites will serve as targets for the effective treatment of obesity. Plateaus in blood pressure occur during treatment with individual drugs and this has led to the use of combination therapy to enhance the desired hypotensive effect. Similarly, plateaus in weight occur with any therapy for obesity as regulatory compensation comes into play. Thus, like hypertension, it is reasonable to expect that for many patients, effective therapy for obesity will involve the use of more than one drug. Finally, treatment of hypertension is chronic because the disease recurs when treatment is stopped. Similarly, for patients with obesity

Box 1

Five strategies for primary drug action that might lead to significant weight loss

1. Reducing food intake either by amplifying inhibitory effects of anorexigenic signals or factors (those that suppress food intake) or by blocking orexigenic signals or factors (those that stimulate food intake).

2. Blocking nutrient absorption (especially fat) in the gut.

3. Increasing thermogenesis by uncoupling of fuel metabolism from the generation of ATP, thereby dissipating food energy as heat (see review by Lowell and Spiegelman, pp. 652–660).

4. Modulating fat or protein metabolism or storage by regulating fat

synthesis/lipolysis or adipose differentiation/apoptosis. Enhanced fat or protein turnover might reduce body weight by affecting either food intake or energy expenditure.

5. Modulating the central controller regulating body weight

(i) by altering the internal reference value sought by the controller or

(ii) by modulating the primary afferent signals regarding fat stores that are analysed by the controller. This approach would have the potential advantage of forcing the endogenous controller to regulate multiple pathways of energy balance and minimize compensation.

who show improvements in their associated metabolic risk factors, treatment may also need to be chronic. When treatment of obesity is primarily for cosmetic purposes, however, the duration of treatment may be shorter or intermittent rather than chronic, thus reducing the risks and cost of chronic exposure to drugs.

When body-weight regulation is viewed as a feedback system, it can be analysed as four separate components: afferent signals such as leptin, which are produced by fat and signal to the brain; a central controller in the brain; efferent signals that modulate food seeking, food ingestion, energy expenditure and metabolism; and the controlled system that digests, absorbs and stores food energy⁵. The key elements of such a system are shown in Fig. 1 (refs 6,7). Potential therapeutic intervention points are located throughout this feedback system. It is important to note that there is not a one-to-one correspondence between the four components of this feedback system and the individual antiobesity strategies (such as reducing food intake or increasing thermogenesis) that are outlined in Box 1 and discussed in more detail below. This is because each of the basic antiobesity drug strategies could be realized through manipulation of multiple points within this feedback system.

For a drug to have a significant impact on body weight it must ultimately affect energy intake, energy expenditure, or both. This is a result of the first law of thermodynamics, which states that the amount of stored energy equals the difference between energy intake and work.

It is possible that drugs of some benefit to the obese could be developed that affect neither energy intake nor expenditure, but rather affect the partitioning of nutrients away from adipose tissue and towards other organ systems. However, owing to the limited storage capacity of other organ systems and the large excess of stored energy typical of the obese, drugs that fail to impact upon energy intake or expenditure are likely to be of limited significance.

Furthermore, drugs that act 'purely' on one end of the energy balance equation (that is, on energy intake or expenditure) may also fail to achieve significant long-term efficacy. Because body fat is likely to be regulated homeostatically, a change in either intake or expenditure alone will face resistance as compensatory adjustments are made. For example, a drug that primarily targets thermogenesis must at least attenuate any compensatory increase in food intake that occurs as a result of the intervention. In the future, it is possible that combined drug therapy will solve this dilemma.

Drugs currently available

Sibutramine — a strategy to reduce food intake

Produced under the registered trademarks Meridia and Reductil by Knoll Pharmaceuticals, Mount Olive, New Jersey, and Ludwigshafen, Germany, sibutramine is a β -phenethylamine that selectively inhibits the reuptake of noradrenaline, serotonin (5-HT) and, to a lesser extent, dopamine. In experimental animals and in human beings, sibutramine reduces food intake⁸ and arguably increases thermogenesis^{9,10}. The effects of sibutramine on food intake are similar to the effects of simultaneous administration of drugs that selectively block reuptake of serotonin (fluoxetine) and noradrenaline (nisoxetine). Thus the main therapeutic reduction of food intake in humans probably results from blocking reuptake of both serotonin and noradrenaline.

There is a dose-related reduction in body weight¹¹ in clinical trials with sibutramine, with weight loss up to 9% below baseline, which can last up to 18 months with continued treatment. When weight loss is induced with diet, patients randomized to the sibutramine treatment continued to lose weight over a one-year period, reaching 15% below baseline, whereas the placebo-treated patients regained some weight¹². In the Sibutramine Trial of Obesity Reduction and Maintenance (STORM Trial), the group randomized to sibutramine after weight loss maintained their lower weight for up to 18 months, whereas the placebo-treated patients regained weight (W. P. T. James, unpublished results presented at the European Congress on Obesity,

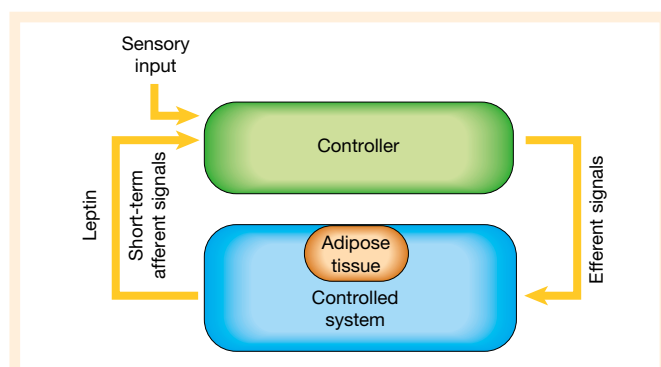


Figure 1 Body-weight regulation can be viewed as a feedback system. Multiple sites in this system may serve as targets for effective treatment of obesity. The brain is the controller that serves as a receiver and transducer for input signals, and an integrator of output signals to the neural and hormonal systems. These efferent output signals modulate food seeking behaviour and regulate metabolism. The controlled system processes, stores and metabolizes the nutrients obtained from food. The afferent signals comprise neural (vagus and sensory) and hormonal (leptin) inputs that tell the brain (controller) about the state of the body and the world outside.

Milan, 1999). The principal side effects are those associated with the sympathomimetic properties of the drug, and include dry mouth, insomnia and asthenia¹¹. A small increase in blood pressure and heart rate also occurs and persists for as long as treatment is continued, which therefore requires monitoring.

Orlistat — a strategy to block fat absorption

Produced under the registered trademark Xenical by Hoffmann-LaRoche, Basel, Switzerland, and Nutley, New Jersey, Orlistat is a hydrogenated derivative of a bacterial lipase inhibitor that blocks pancreatic lipase, thus decreasing triglyceride digestion¹³. In dose-ranging studies on faecal fat loss, orlistat blocked digestion of up to 30% of orally ingested triglyceride on a diet containing 30% fat at a dose of ~120 mg three times daily, the dose that is now used clinically¹⁴. Two published clinical trials^{15,16}, each lasting two years, have shown that after one year the drug produces a weight loss of about 9–10% compared with a 4–6% weight loss in a placebo-treated group. When the subjects were re-randomized at one year to orlistat or placebo, the patients treated with placebo in the first year who received drug in the second year lost weight, whereas those switched from orlistat to placebo gained weight. In one trial in diabetics the weight loss was smaller, reaching 6% after one year compared with 4% for placebo-treated patients¹⁷. In a secondary prevention study for weight maintenance, patients were randomized to orlistat or placebo after dieting to lose at least 8% of baseline weight; those randomized to orlistat regained less weight (32.4%) after 12 months than those randomized to placebo (56% regain). Because orlistat blocks the action of lipase, faecal fat loss increases¹⁸ and may be associated with gastrointestinal side effects such as oily spotting, faecal urgency and increased defecation. These side effects decline over time¹⁹ as patients learn how to use the medication. Absorption of fat-soluble vitamins, particularly vitamin E and β -carotene, may be reduced, but plasma vitamin levels usually remain within normal limits. Although gastrointestinal symptoms may be a reason for discontinuation of the drug, it seems to be a minor problem for most patients.

Ephedrine and caffeine — a strategy to increase energy expenditure

Ephedrine is thermogenic²⁰: when administered to human subjects it increases oxygen consumption about 10% over a period of several hours, and the effect is dose-related²⁰. In clinical trials, ephedrine and caffeine produced significantly more weight loss than placebo, ephedrine, or caffeine alone^{21,22}. The major side effects are an increase in heart rate and a sense of palpitations in some patients. Of interest, this combination reduces the loss of lean body mass while enhancing the oxidation of fatty acids, and thus might be classified as a 'nutrient

partitioning' agent²³ (see below). It is estimated that only 25 to 40% of the weight loss in patients treated with ephedrine and caffeine is due to thermogenesis; the remaining 60–75% coming from a reduction in food intake²³. Thus, the one 'thermogenic' drug combination tested clinically had more effect on reducing food intake than on stimulating energy expenditure²³.

Strategies for new drug discovery

Research over the past few years has provided an unprecedented expansion of our knowledge about the molecular mechanisms regulating body fat. Perhaps the greatest impact has resulted from the cloning of genes corresponding to the five monogenic obesity syndromes in mice and the subsequent characterization of the pathways identified by these genetic entry points. Three of these genes (*ob*, *db* and *A'*) have already led to potential drugs or drug targets that are currently in pharmaceutical development: respectively, these are leptin, leptin receptor and melanocortin (MC)-4 receptor^{24–29} (see review by Barsh *et al.*, pp. 644–651). In addition, extensive molecular and reverse genetic studies (mouse knockouts) have helped establish other critical factors in energy balance, as well as validated or refuted the importance or previously identified pathways. This new information, combined with the increased recognition of obesity as a serious disease, has encouraged obesity drug development across the pharmaceutical industry. Potential strategies and specific targets for antiobesity drugs that have resulted from these mechanistic insights are discussed below (Fig. 2).

Targeting molecules that regulate food intake peripherally

Gastrointestinal peptides have long been studied as potential regulators of satiety. Cholecystokinin (CCK) was one of the first peptides shown to reduce food intake. This occurs in animals and human beings alike³⁰. Peptide analogues of CCK have been developed, but none has reached the clinic, suggesting that they may have undesirable side effects. Gastrin-releasing peptide, neuromedin B, and bombesin, the peptide derived from frog skin, reduce food intake in animals and human beings³¹. If the duration of action of these peptides or an effective analogue could be developed it would have a potential place in modulating intake at individual meals.

Pancreatic peptides modulate feeding. Both glucagon and glucagon-like peptide-1 (GLP-1), a derived peptide of 6–29 amino acids, reduce food intake in animals and humans^{32–34}. Analogues or small molecules that might influence GLP-1 receptors, GLP-1 release or the duration of action would be interesting candidates for development. Enterostatin, the pentapeptide signal portion of pancreatic co-lipase, is of interest because it selectively reduces fat intake in experimental animals³⁵. This peptide increases satiety in humans³⁶ and reduces food intake in baboons³⁷. Once again, a small-molecule analogue could be useful as a strategy for reducing intake of single meals. Finally, amylin may have effects of food intake³⁸.

Nutrient signals provide molecules that modulate feeding and may also serve as models for antiobesity drugs. In experimental animals, pyruvate, lactate and 3-hydroxy-butyrate reduce food intake^{39,40}. So, too, do analogues of these molecules⁶. Hydroxycitrate, a molecule that can inhibit ATP-citrate lyase, was shown to reduce obesity in experimental animals⁴¹, but clinical trials showed no effect⁴². Amino acids might also be a target. Selective deficiency of individual amino acids reduces food intake and this may involve 5-HT₃ receptors in the pre-pyiform cortex⁴³. Part of the ApoA-IV molecule, which is produced by the intestine and involved in the transport of fat from the intestine to the circulation, has been found to reduce food intake⁴⁴ and might be a target for drug development.

Targeting molecules that regulate food intake centrally

In recent years, the most intensively explored areas of food-intake regulation have been analyses of central neuropeptide and monoamine mechanisms of action. Leptin is the best known of the afferent fat signals and the best candidate for the primary signal

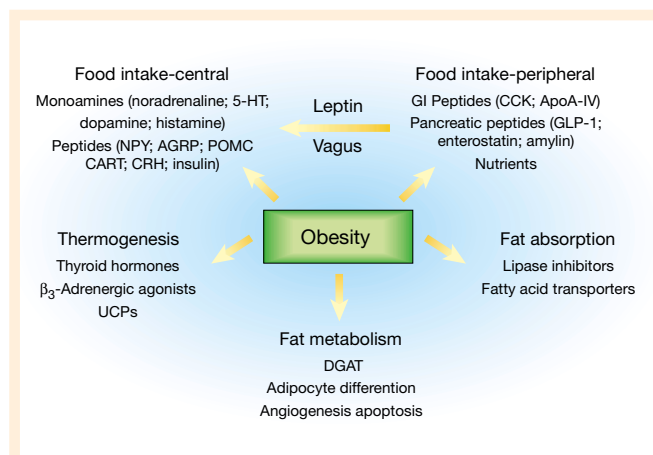


Figure 2 Diagram of strategies for molecules targeted against obesity.

communicating body fat information to the central controller^{24,45}. Identification of this peptide through positional cloning in 1994 provided major new insights into the regulation of food intake, energy expenditure and body fat. It is now clear that this cytokine, derived primarily from fat cells, but also from the placenta and possibly the stomach, reduces food intake and increases the activity of the thermogenic components of the sympathetic nervous system. Modulation of neurons in the arcuate nucleus by leptin results in reduced secretion of neuropeptide Y (NPY), reduced expression of agouti-related protein (AGRP), and increased expression of pro-opiomelanocortin (POMC; the precursor of α -melanocyte-stimulating hormone (α -MSH) that reduces food intake) and the peptide product of cocaine- and amphetamine-regulated transcript (CART)^{46–48}.

Because of these coordinated effects of leptin, this adipocyte-derived cytokine has been used in a clinical trial where there was a dose-related weight loss, but with significant discomfort at the injection site⁴⁹. It has also been shown to be clinically effective in those rare individuals who lack leptin⁵⁰ and to eliminate almost completely body fat in a transgenic mouse that overexpresses leptin⁵¹. Despite the limited effect of leptin in this early trial⁴⁹, there is still great potential for a leptin-like product or a drug that can activate the intracellular signalling system modulated by leptin. The development of small-molecule agonists of the leptin receptor remains a viable strategy. These could be both orally available and penetrate the blood–brain barrier, thereby allowing much more potent stimulation of the leptin receptor than a protein compound limited by saturable transport across the blood–brain barrier. Alternatively, sensitizers of the leptin pathway could be developed, as is now occurring with insulin sensitizers in type II diabetes. In addition, ciliary neurotrophic factor (CNTF) is a neurocytokine that is expressed in glial cells and reduces food intake and body weight in rodents, possibly through the activation of the janus kinase/signal transducer and activator of transcription (JAK/STAT) pathway in leptin receptor-expressing cells⁵². Although CNTF is unlikely to be involved in normal body-weight regulation, its potential pharmacological value makes this cytokine worth investigating.

The number of neuropeptides shown to affect food intake has been growing in recent years at an accelerating rate. However, the degree of biological validation and probable relative importance varies considerably. Neuropeptide Y is among the most potent stimulators of feeding⁵³; its synthesis and release is modulated by insulin, leptin and starvation. Antagonists to either the Y-5 or the Y-1 NPY receptor are being explored as potential agents for treatment of obesity. Knockout of the NPY, NPY-Y1 and NPY-Y5 receptor does not affect nutrient status⁵⁴, which implies that redundant systems may replace NPY when it is absent. This may also limit the use of potential

NPY antagonists.

The melanocortin-receptor system is an attractive target for therapeutic advance⁴⁶, particularly since genetic defects in the human POMC and MC4 receptor genes have been reported in extremely obese individuals. A natural agonist, α -MSH, reduces food intake, and a mouse lacking POMC, the precursor of α -MSH, is obese⁵⁵. The biological importance of this receptor system became clear when it was shown that MC4-receptor-knockout mice became massively obese, approaching the weight of a leptin-deficient obese mouse²⁹. This indicates that specific MC4-receptor agonists may become important obesity therapeutics. The melanocortin-receptor system is the one on which the genetic defect in the agouti locus of the yellow mouse plays out its role (see review by Barsh *et al.*, pp. 644–651).

Antagonists to melanin-concentrating hormone (MCH) receptor are another potential approach for drug development^{56–59}. MCH is produced by neurons in the lateral hypothalamus and microinjection of this peptide increases food intake. MCH-knockout mice are lean, suggesting that the peptide has a physiological role in the control of food intake and body fat stores. The MCH receptor has recently been identified⁵⁹ and is undoubtedly being used to screen for MCH antagonists.

The opioid receptors were the first group of peptide receptors shown to modulate feeding⁶⁰. Both the mu and kappa receptor can affect feeding and macronutrient selection by modulating fat intake, but whether this strategy can be applied to development of new drugs because of the linguistic connotations of 'opioid-like' is uncertain.

Several other peptides that stimulate food intake are of lesser interest to us because they have been associated with other significant biological events. Galanin is an endogenous peptide that will increase food intake when injected into the brain's ventricular system. Mice lacking galanin, however, are unable to maintain lactation, suggesting that modulation of milk-producing hormones may be the primary role for this peptide⁶¹. Orexin A was identified originally as a peptide that stimulates food intake⁶², but defects in the orexin peptides and receptors causes narcolepsy in mice and dogs^{63,64}. Thus the orexin peptide, which is abundant in the lateral hypothalamus, serves an arousal function, and its effects on food

intake may be secondary. Whether sleepiness is a good alternative to over-eating is debatable. Corticotropin-releasing hormone (CRH) and the closely related urocortin affect food intake and body weight⁶⁵, and CRH receptors and CRH binding protein have therefore been considered as antiobesity targets. However, intervention with this pathway may have negative consequences on the stress axis and anxiety⁶⁶.

Table 1 includes several additional peptides that have been shown to reduce food intake. The list of inhibitory peptides is longer than the list that stimulates food intake. However, some of these peptides may reduce food intake by producing nausea or sedation.

The serotonin system has been the most heavily studied of the monoamine pathways and serotonin receptors modulate both the quantity of food and macronutrient selection. Stimulation of the serotonin receptors in the paraventricular nucleus reduces fat intake with little or no effect on the intake of protein or carbohydrate. The reduction in fat intake by fenfluramine is probably through 5-HT_{2C} receptors, as its effect is attenuated in 5-HT_{2C}-knockout mice⁶⁷. Accordingly, specific 5-HT_{2C} agonists may be an important future therapeutic target, although the possibility that desensitization of these receptors might increase seizure susceptibility (a feature noted in the knockout mice) should be kept in mind. Although serotonin receptors may be effective sites for modulation of both total intake and fat intake⁶⁸, the unfortunate development of valvulopathy with fenfluramine and dexfenfluramine will surely limit further efforts at new drug development using serotonin receptors^{69–71}.

Noradrenergic receptors are also involved in modulating food intake. The stimulation of α_1 -noradrenergic receptors reduces food intake. Phentolamine is an α_1 -agonist that has a modest effect on food intake. Some of the antagonists to the α_1 -receptors that are used to treat benign prostatic hypertrophy produce weight gain, indicating that this receptor is clinically important. Stimulation of α_2 -receptors will increase food intake in experimental animals and a polymorphism in the α_{2B} -adrenoceptor has been associated with reduced metabolic rate in humans⁷². On the other hand, the β_2 -receptors in the brain reduce food intake. Agonist drugs can activate these receptors by releasing noradrenaline in their vicinity, or by blocking the reuptake of noradrenaline.

Stimulation of dopamine receptors reduces food intake⁷³. Both the D1/D5 and the D2/D3/D4 receptors seem to have this effect. However, they also influence the preference for foods and the hedonic experiences surrounding food. Whether they offer effective approaches to treatment of obesity remains to be determined.

Pharmacologically defined central histamine H₃ receptors have also been identified experimentally as modulators of feeding⁷⁴. Recently, the histamine H₃-receptor molecule has been cloned⁷⁵, which should enable more detailed studies of the role of H₃ receptors in body-weight regulation. It is conceivable that some of the weight gain seen with the broad-based phenothiazines may result from their interaction with histaminergic receptors. A recent paper indicates that the effects of histamine may be through its effect on release of noradrenaline⁷⁶.

Targeting molecules in fat absorption

The successful introduction of Xenical to the market has demonstrated that drugs that inhibit the absorption of dietary fat can be an important component of antiobesity therapy. However, side effects of Xenical resulting from faecal fat loss can be a problem for some patients. Another approach to inhibiting the absorption of dietary fat may be to allow the digestion of fat to fatty acids in the gut, yet block fatty acid uptake by the intestine. The recent identification of the major intestinal fatty acid transporter (FATP4)⁷⁷ may enable the identification of drugs that inhibit this transport process. Such drugs may have the potential advantage of preventing side effects that are associated with fatty stools, but it may also enhance the loss of divalent cations. Thus, the

Table 1 Centrally acting monoamine and neuropeptide systems that suppress food intake

Target system	Potential therapeutic strategies
Noradrenergic	α_1 -receptor agonists
	β_2 -receptor agonists
	Stimulate noradrenaline (NE) release
	Block NE uptake
Serotonergic	Stimulate 5-HT release
	Block 5-HT uptake
	5-HT _{2C} -receptor agonists
Dopaminergic	D1-receptor agonists
Histaminergic	H3-receptor antagonists
Leptin	Leptin/leptin analogues
	Leptin-receptor (Ob-R) agonists
	Leptin sensitizers
Melanocortin	MC4-receptor agonists
NPY	NPY-Y1 or -Y5 antagonists
MCH	MCH-receptor antagonists
CRH/urocortin	Receptor agonists
	CRH-BP blockers
Galanin	Receptor antagonists
Orexin	Receptor antagonists
Opioid	Mu and kappa receptors
GART	Receptor agonists
Apo (A ^{IV})	Receptor agonists
Amylin	Receptor agonists

consequences of unabsorbed fatty acids in the stool must be carefully evaluated.

Targeting thermogenesis

Stimulation of thermogenesis has been used to treat obesity for more than 100 years, when thyroid extract was first used^{5,6}. Although thyroid hormones increase energy expenditure and reduce body fat, they also increase the loss of lean body mass and bone calcium. If these effects could be eliminated, the thermogenic effect and loss of fat might be beneficial. The use of dinitrophenol, a thermogenic drug that uncouples oxidative phosphorylation, produced weight loss but had the highly undesirable side effect of producing cataracts and neuropathy⁷⁸.

β_3 -Adrenergic-receptor agonists are currently being evaluated in clinical programmes in several pharmaceutical companies. Such drugs have the potential of increasing energy expenditure⁷⁹ (see review by Kopelman, pp. 635–643). Previous versions of β_3 -agonists were unsuccessful in clinical trials, but this was due to lack of specificity and potency for the human β_3 -receptor, and to the hand tremor and increased tachycardia seen with early drugs. Within the next few years, the improved β_3 -agonists currently being developed should address whether or not this is a viable strategy.

Mitochondria in brown adipose tissue contain an uncoupling protein (UCP-1) that has a well-established role in uncoupling oxidative metabolism from ATP generation. UCP-1 is involved in rodent temperature and body-weight regulation (see review by Lowell and Spiegelman, pp. 652–660). Increased expression and/or activation of this protein uncouples oxidative phosphorylation resulting in the conversion of energy to heat (thermogenesis)⁸⁰. The importance of this molecule in humans has always been questioned owing to the low levels of brown fat (and hence UCP-1 expression) in adult humans. But the identification of two additional uncoupling proteins (UCP-2 and UCP-3) that are highly expressed in adult human tissues has attracted considerable interest^{81–85}. It is possible that drugs activating or increasing the expression of UCP-2 and UCP-3 would have important effects on energy expenditure or on the rate of nutrient oxidation.

The enzyme protein tyrosine phosphatase-1B (PTP-1B) has been implicated in insulin resistance. Recently it has been shown that insulin resistance in PTP-1B-knockout mice is reduced, yet the mice appear otherwise healthy⁸⁶. Of interest for obesity is the finding that these mice do not become obese when eating a high-fat diet. The reason for the protection from diet-induced obesity is currently unknown, but these data indicate that drugs that inhibit PTP-1B activity may have potential as antiobesity agents.

Targeting fat metabolism

Enzymes involved in fat metabolism may be important obesity targets. A notable example is the recently cloned acyl CoA:diacylglycerol transferase (DGAT) activity⁸⁷, the enzyme responsible for the final step in the glycerol phosphate pathway of triglyceride synthesis. Proteins involved in adipocyte differentiation, angiogenesis or apoptosis could also be targeted as a way to reduce fat mass. However, for each of these potential targets it must be shown that the reduced ability to synthesize or store fat is sensed by the body and translated into decreased energy intake or increased energy expenditure. Otherwise the inability to deliver excess calories to adipose tissue could have serious secondary consequences as lipids accumulate in the blood or various organ systems.

The hormonal systems that modulate cellular metabolism are also of interest. During growth, growth hormone and thyroid hormone work together to increase growth of the body. At puberty, newly produced gonadal steroids lead to shifts in the relationship of fat to lean body mass in boys and girls. Testosterone increases lean mass relative to fat and oestrogen has the opposite effect. Testosterone levels fall when human males grow older, and there is a corresponding increase in visceral and total body fat and a decrease in

lean body mass. This may be compounded by the decline in growth hormone that is also associated with an increase in fat relative to lean mass. Both testosterone and growth hormone have been used to treat obesity^{6,88–91}. Growth hormone increases energy expenditure and increases the loss of fat^{6,92}, whereas testosterone and anabolic steroids in males can lower visceral fat relative to total body fat, suggesting selective effects on different fat deposits^{88–91}. Growth hormone also reduces visceral fat more than total fat. Both testosterone and growth hormone have undesirable side effects — testosterone is related to the development of prostatic hypertrophy and prostatic cancer, whereas excess growth hormone has been associated with enhanced risk of cardiovascular disease. The local injection or topical application of lipolytic drugs has also been reported to have modest effects in reducing local fat deposits⁹³, and this cosmetic effect could be particularly appealing to some individuals.

Conclusions

Current drugs for treatment of obesity have a useful place as part of the treatment programme for some overweight patients. But because obesity is a worldwide epidemic, there is a major need for more medicinal products that are proven safe and effective when used as prescribed. In addition, as obesity is associated frequently with other disorders such as diabetes, hypertension, dyslipidaemia, sleep apnoea and osteoarthritis, it is important that the physician treat all of the relevant treatable diseases. For the scientist, the challenge of obesity is to identify those major pathways where pharmacological intervention can safely and effectively help those who are clinically overweight. For the public at large, efforts at prevention of obesity are long overdue to stem the potentially devastating consequences of future associated diseases such as diabetes and its attendant problems. □

1. World Health Organization. *Obesity: Preventing and Managing the Global Epidemic*. (World Health Organization, Geneva, 1998).
2. National Heart Lung and Blood Institute. Clinical guidelines on the identification evaluation and treatment of overweight and obesity in adults — the evidence report. *Obes. Res.* **6**(Suppl. 2), 51S–290S (1998).
3. Allison D. B. *et al.* Annual deaths attributable to obesity in the United States. *J. Am. Med. Assoc.* **282**, 1530–1538 (1999).
4. Bray, G. A. Drug treatment of obesity. Don't throw out the baby with the bath water. *Am. J. Clin. Nutr.* **67**, 1–2 (1998).
5. Bray, G. A. *Current and Contemporary Management of Obesity* (Handbooks in Health Care, Newtown, PA, 1998).
6. Bray, G. A. & Greenway, F. L. A review of current and potential drugs for treatment of obesity. *Endocr. Rev.* **20**, 805–875 (1999).
7. Jequier, E. & Tappy, L. Regulation of body weight in humans. *Physiol. Rev.* **79**, 451–480 (1999).
8. Rolls, B. J., Shide, D. J., Thorwart, M. L. & Ulbrecht, J. S. Sibutramine reduces food intake in non-dieting women with obesity. *Obes. Res.* **6**, 1–11 (1998).
9. Hansen, D. L. *et al.* Thermogenic effects of sibutramine in humans. *Am. J. Clin. Nutr.* **68**, 1180–1186 (1998).
10. Seagle, H. M., Gessesen, D. H. & Hill, J. O. Effects of sibutramine on resting metabolic rate and weight loss in overweight women. *Obes. Res.* **6**, 115–121 (1998).
11. Bray, G. A. *et al.* Sibutramine produces dose-related weight loss. *Obes. Res.* **7**, 189–198 (1999).
12. Apfelbaum, M. *et al.* Long-term maintenance of weight loss after a very low calorie diet: efficacy and tolerability of sibutramine. *Am. J. Med.* **106**, 179–184 (1999).
13. Guercolini, R. Mode of action of orlistat. *Int. J. Obes. Relat. Metab. Disord.* **21**, S12–S23 (1997).
14. Hauptman, J. B., Jeunet, F. S. & Hartmann, D. Initial studies in humans with the novel gastrointestinal lipase inhibitor Ro 18-0647 (tetrahydrolipstatin). *Am. J. Clin. Nutr.* **55**, 309S–313S (1992).
15. Sjostrom, L. *et al.* Randomized placebo-controlled trial of orlistat for weight loss and prevention of weight regain in obese patients. European Multicentre Orlistat Study Group. *Lancet* **352**, 167–172 (1998).
16. Davidson, M. H., Hauptman, J. & DiGirolamo, M. Weight control and risk factor reduction in obese subjects treated for 2 years with orlistat. A randomized controlled trial. *J. Am. Med. Assoc.* **281**, 235–242 (1999).
17. Hollander, P. *et al.* Role of orlistat in the treatment of obese patients with type 2 diabetes. *Diabetes Care* **21**, 1288–1294 (1998).
18. Hill, J. O. *et al.* Orlistat, a lipase inhibitor, for weight maintenance after conventional dieting — A 1 year study. *Am. J. Clin. Nutr.* **9**, 1108–1116 (1999).
19. Conavatchel, W. Long-term tolerability profile of orlistat, an intestinal lipase inhibitor. *Diabetologia* **40**, A196 (1997).
20. Astrup, A., Lundsgaard, C., Madsen, J., & Christensen, N. J. Enhanced thermogenic responsiveness during chronic ephedrine treatment in man. *Am. J. Clin. Nutr.* **42**, 83–94 (1985).
21. Astrup, A., Breum, L., Toubro, S., Hein, P. & Quaade, F. The effect and safety of an ephedrine/caffeine compound compared to ephedrine, caffeine and placebo in obese subjects on an energy-restricted diet. A double-blind trial. *Int. J. Obes. Relat. Metab. Disord.* **16**, 269–277 (1992).
22. Toubro, S., Astrup, L., Breum, L. & Quaade, F. The acute and chronic effects of ephedrine/caffeine mixtures on energy expenditure and glucose metabolism in humans. *Int. J. Obes. Relat. Metab.*

- Disord.* **17**, S73–S77 (1993).
23. Astrup, A. *et al.* The effect of ephedrine/caffeine mixture on energy expenditure and body composition in obese women. *Metabolism* **41**, 686–688 (1992).
24. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
25. Tartaglia, L. A. *et al.* Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**, 1263–1271 (1995).
26. Chen, H. *et al.* Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in db/db mice. *Cell* **84**, 491–495 (1996).
27. Lee, G. H. *et al.* Abnormal splicing of the leptin receptor in diabetic mice. *Nature* **379**, 632–635 (1996).
28. Chua, S. C. Jr *et al.* Phenotypes of mouse diabetes and rat fat due to mutations in the OB (leptin) receptor. *Science* **271**, 994–996 (1996).
29. Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131–141 (1997).
30. Smith, G. P. *Satiation: From Gut to Brain* 291 (Oxford Univ. Press, New York, 1998).
31. Gutzwiller, J. P. *et al.* Effect of intravenous human gastrin-releasing peptide on food intake in humans. *Gastroenterology* **106**, 1168–1173 (1994).
32. Flint, A., Raben, A. L., Astrup, A. & Holst, J. J. Glucagon-like peptide I promotes satiety and suppresses energy intake in humans. *J. Clin. Invest.* **101**, 515–520 (1998).
33. Holst, J. J. Glucagon-like peptide-1 (GLP-1) — an intestinal hormone, signaling nutritional abundance, with an unusual therapeutic potential. *Trends Endocrinol. Metab.* **10**, 229–235 (1999).
34. Gutzwiller, J. P. *et al.* Glucagon-like peptide-1 promotes satiety and reduces food intake in patients with diabetes mellitus type 2. *Am. J. Physiol.* **276**(5 Pt 2), R1541–R1544 (1999).
35. Erlanson-Albertsson, C. & York, D. Enterostatin — a peptide regulating fat intake. *Obes. Res.* **5**, 360–372 (1997).
36. Smeets, M., Geiselman, P., Bray, G. A. & York, D. A. The effect of oral enterostatin on hunger and food intake in human volunteers. *FASEB J.* **13**, A871 (1999).
37. Wetherford, S. C. *et al.* Intraventricular administration of enterostatin decreases food intake in baboons. *Appetite* **19**, 225 (1992).
38. Morley, J. E., Flood, J. F., Horowitz, M., Morley, P. M. & Walter, M. J. Modulation of food intake by peripherally administered amylin. *Am. J. Physiol.* **276**(1 Pt 2), R178–R184 (1994).
39. Nagase, H., Bray, G. A. & York, D. A. Effects of pyruvate and lactate on food intake in rat strains sensitive and resistant to dietary obesity. *Physiol. Behav.* **59**, 555–560 (1996).
40. Scharrer, E. Control of food intake by fatty acid oxidation and ketogenesis. *Nutrition* **15**, 704–714 (1999).
41. Sullivan, A. C., Triscari, J., Hamilton, J. G. & Miller, O. N. Effect of (–)-hydroxycitrate upon the accumulation of lipid in the rat. II. Appetite. *Lipids* **9**, 129–134 (1974).
42. Heymsfield, S. B. *et al.* Garcinia cambogia (hydroxycitric acid) as a potential antiobesity agent: a randomized controlled trial. *J. Am. Med. Assoc.* **280**, 1596–1600 (1998).
43. Gietzen, D. W., Erecius, L. F. & Rogers, Q. R. Neurochemical changes after imbalanced diets suggest a brain circuit mediating anorectic responses to amino acid deficiency in rats. *J. Nutr.* **128**, 771–781 (1998).
44. Tso, P., Liu, M., & Kalogeris, T. J. The role of apolipoprotein A-IV in food intake regulation. *J. Nutr.* **8**, 1503–1506 (1999).
45. Mantzoros, C. S. The role of leptin in human obesity and disease — a review of current evidence. *Ann. Intern. Med.* **130**, 671–680 (1999).
46. Cone, R. D. The central melanocortin system and energy homeostasis. *Trends Endocrinol. Metab.* **10**, 211–216 (1999).
47. Wilson, B. D., Ollmann, M. M. & Barsh, G. S. The role of agouti-related protein in regulating body-weight. *Mol. Med. Today* **5**, 250–256 (1999).
48. Kristensen, P. *et al.* Hypothalamic CART is a new anorectic peptide regulated by leptin. *Nature* **393**, 72–76 (1998).
49. Heymsfield, S. B. *et al.* Recombinant leptin for weight loss in obese and lean adults: a randomized, controlled, dose-escalation study. *J. Am. Med. Assoc.* **282**, 1568–1575.
50. Farooqi, I. S. *et al.* Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *N. Engl. J. Med.* **341**, 879–884 (1999).
51. Ogawa, Y. *et al.* Increased glucose metabolism and insulin sensitivity in transgenic skinny mice overexpressing leptin. *Diabetes* **48**, 1822–1829 (1999).
52. Bjorbaek, C. *et al.* Activation of SOCS-3 messenger ribonucleic acid in the hypothalamus by ciliary neurotrophic factor. *Endocrinology* **140**, 2035–2043 (1999).
53. Zimanyi, I. A., Fathi, Z. & Poindexter, G. S. Central control of feeding behavior by neuropeptide Y. *Curr. Pharm. Des.* **4**, 349–366 (1998).
54. Palmiter, R. D. *et al.* in *Pennington Center Nutrition Series: Nutrition, Genetics, and Obesity* 269–286 (Louisiana State Univ. Press, Baton Rouge, 1999).
55. Yaswen, L., Diehl, N., Brennan, M. B. & Hochgeschwender, U. Obesity in the mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin. *Nature Med.* **5**, 1066–1070 (1999).
56. Chambers, J. *et al.* Melanin-concentrating hormone is the cognate ligand for the orphan G-protein-coupled receptor SLC-1. *Nature* **400**, 261–269 (1999).
57. Saito, Y. *et al.* Molecular characterization of the melanin-concentrating-hormone receptor. *Nature* **400**, 265–269 (1999).
58. Shimada, M. *et al.* Mice lacking melanin-concentrating hormone are hypophagic and lean. *Nature* **396**, 670–673 (1998).
59. Lembo, P. M. *et al.* The receptor for the orexigenic peptide melanin-concentrating hormone is a G-protein-coupled receptor. *Nature Cell Biol.* **5**, 267–271 (1999).
60. Barton, C., York, D. A. & Bray, G. A. Opioid receptor subtype control of galinin-induced feeding. *Peptides* **17**, 237–240 (1996).
61. Rokaeus, A., Jiang, K., Spyrou, G. & Waschek, J. A. Transcriptional control of the galinin gene. Tissue-specific expression and induction by NGF, protein kinase C, and estrogen. *Ann. NY Acad. Sci.* **863**, 1–13 (1998).
62. Sakurai, T. *et al.* Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* **92**, 573–585 (1998).
63. Lin, L. *et al.* The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (Orexin) Receptor 2 gene. *Cell* **98**, 365–376 (1999).
64. Chemelli, R. M. *et al.* Narcolepsy in orexin knockout mice: molecular genetics of sleep regulation. *Cell* **98**, 437–451 (1999).
65. Kalra, S. P. *et al.* Interacting appetite-regulating pathways in the hypothalamic regulation of body weight. *Endocr. Rev.* **1**, 68–100 (1999).
66. Karolyi, I. J. *et al.* Altered anxiety and weight gain in corticotropin-releasing-hormone-binding protein-deficient mice. *Proc. Natl Acad. Sci. USA* **96**, 11595–11600 (1999).
67. Vickers, S. P., Clifton, P. G., Dourish, C. T. & Tecott, L. H. Reduced satiating effect of d-fenfluramine in serotonin 5-HT (2C) receptor mutant mice. *Psychopharmacology* **143**, 309–314 (1999).
68. Smith, B. K., York, D. A. & Bray, G. A. Activation of hypothalamic serotonin receptors reduced intake of dietary fat and protein but not carbohydrate. *Am. J. Physiol.* **277**, R802–R811 (1999).
69. Connolly, H. M. *et al.* Valvular heart disease associated with fenfluramine-phentermine. *N. Engl. J. Med.* **337**, 581–588 (1997).
70. Ryan, D. H. *et al.* Serial echocardiographic and clinical evaluation of valvular regurgitation before, during, and after treatment with fenfluramine or dexfenfluramine and mazindol or phentermine. *Obes. Res.* **7**, 313–322 (1999).
71. Rothman, R. B., Ayestas, M. A., Dersch, C. M., & Baumann, M. H. Aminorex, fenfluramine, and chlorphentermine are serotonin transporter substrates. Implications for primary pulmonary hypertension. *Circulation* **100**, 869–875 (1999).
72. Heinonen, P. *et al.* Identification of a three-amino acid deletion in the alpha2B-adrenergic receptor that is associated with reduced basal metabolic rate in obese subjects. *J. Clin. Endocrinol. Metab.* **84**, 2429–2433 (1999).
73. Terry, P., Gilbert, D. B. & Cooper, S. J. Dopamine receptor subtype agonists and feeding behavior. *Obes. Res.* **3**, 515S (1995).
74. Sakata, S., Yoshimatsu, H. & Kurokawa, M. Hypothalamic neuronal histamine: implications of its homeostatic control of energy metabolism. *Nutrition* **13**, 403–411 (1997).
75. Lovenberg, T. W. *et al.* Cloning and functional expression of the human histamine H3 receptor. *Mol. Pharmacol.* **55**, 1101–1107 (1999).
76. Kurose, Y. & Terashima, Y. Histamine regulates food intake through modulating noradrenaline release in the para-ventricular nucleus. *Brain Res.* **15**, 115–118 (1999).
77. Stahl, A. *et al.* Identification of the major intestinal fatty acid transport protein. *Mol. Cell* **4**, 299–308 (1999).
78. Bray, G. A. *The Obese Patient* (Saunders, Philadelphia, 1976).
79. Danforth, E. Jr & Himms-Hagen, J. H. Obesity and diabetes and the beta-3 adrenergic receptor. *Eur. J. Endocrinol.* **136**, 362–365 (1997).
80. Klengenberg, M. & Huang, S. G. Structure and function of the uncoupling protein from brown adipose tissue. *Biochim. Biophys. Acta* **8**, 271–296 (1999).
81. Fleury, C. *et al.* Uncoupling protein-2: a novel gene linked to obesity and hyperinsulinemia. *Nature Genet.* **15**, 269–272 (1997).
82. Gimeno, R. E. *et al.* Cloning and characterization of an uncoupling protein homolog: a potential molecular mediator of human thermogenesis. *Diabetes* **46**, 900–906 (1997).
83. Vidal-Puig, A., Solanes, G., Grujic, D., Flier, J. S. & Lowell, B. B. UCP3: an uncoupling protein homologue expressed preferentially and abundantly in skeletal muscle and brown adipose tissue. *Biochem. Biophys. Res. Commun.* **235**, 79–82 (1997).
84. Boss, O. *et al.* Uncoupling protein-3: a new member of the mitochondrial carrier family with tissue-specific expression. *FEBS Lett.* **408**, 39–42 (1997).
85. Gong, D. W., He, Y., Karas, M. & Reitman, M. Uncoupling protein-3 is a mediator of thermogenesis regulated by thyroid hormone, beta3-adrenergic agonists, and leptin. *J. Biol. Chem.* **39**, 24129–24132 (1997).
86. Elchebly, M. *et al.* Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* **283**, 1544–1548 (1999).
87. Cases, S. *et al.* Identification of a gene encoding an acyl CoA: diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proc. Natl Acad. Sci. USA* **95**, 13018–13023 (1998).
88. Marin, P. *et al.* Androgen treatment of abdominally obese men. *Obes. Res.* **1**, 245–251 (1993).
89. Marin, P. *et al.* The effects of testosterone on body composition and metabolism in middle-aged obese men. *Int. J. Obes. Relat. Metab. Disord.* **16**, 991–997 (1992).
90. Lovejoy, J. C. *et al.* Oral anabolic steroid treatment, but not parenteral androgen treatment, decreases abdominal fat in obese, older men. *Int. J. Obes. Relat. Metab. Disord.* **19**, 614–624 (1995).
91. Kin, K. R. *et al.* Low-dose growth hormone treatment with diet restriction accelerates body fat loss, exerts anabolic effect and improves growth hormone secretory dysfunction in obese adults. *Hormone Res.* **51**, 78–84 (1999).
92. Bujalska, I. J., Kumar, S., Hewison, M. & Stewart, P. M. Differentiation of adipose stromal cells: the roles of glucocorticoids and 11 beta-hydroxysteroid dehydrogenase. *Endocrinology* **140**, 3188–3196 (1999).
93. Greenway, F. L. & Bray, G. A. Topical fat reduction. *Obes. Res.* **3**(Suppl. 4), 561S–568S (1995).



Roche and Obesity

This *Nature* supplement comes at a time when obesity is reaching staggering proportions not only in the United States, but also around the world. In the U.S. alone, obesity is a leading cause of disability and death, second only to smoking. More than half of all Americans are overweight and one in three is obese (defined as having a Body Mass Index [BMI] above 30). Excess weight increases the risk of other severe health problems, including hypertension, hypercholesterolemia, heart disease, diabetes, and some cancers. In 1995, the direct obesity-related health care costs in the U.S. exceeded \$50 billion — nearly 6% of the entire national expenditure for health care. Yet, the prevailing sentiment (including that of many health care providers) labels obesity as a cosmetic issue caused by lack of willpower on the part of obese individuals.

Fortunately, today we now know more about the causes of obesity than ever before. The last decade has shed much light on how body weight is controlled. A major breakthrough was the cloning of the first obesity gene, *ob*. This discovery opened the door to a far greater understanding that obesity is a complex disease of appetite regulation and energy metabolism, which is controlled by many factors. In attempting to understand the underlying cause of obesity and diabetes Roche, in collaboration with partners such as Millennium Pharmaceuticals, Inc., has pursued a genetic approach. This approach, coupled with a major investment in this area, has resulted in the identification of several genes associated with obesity and diabetes in the *ob* pathway related to energy expenditure and satiety. This research has also given us tremendous opportunities to identify novel molecular targets for effective drug therapies.

However, a disease that is so complex — and increasingly so culturally insidious — demands a varied approach to finding successful treatments. The following articles give a sense of the multi-faceted research being done to address this important disease and its associated risk factors. Our collective research findings are critical to changing the prevailing medical and public perception about obesity and recognizing it as a serious disease with grave complications.

We now know that a sustained loss of as little as 5% to 10% of body weight has been shown to improve co-morbidities associated with obesity such as diabetes and hypertension. Weight loss has also been shown to lower levels of triglycerides, decrease total cholesterol and increase levels of HDL cholesterol. Another important factor is that it can lead to improvement in other obesity-related conditions such as osteoarthritis, sleep apnea and pulmonary and cardiac dysfunction.

There is no doubt that, in addition to lifestyle and behavior changes, innovative new drugs will play an important role in managing obesity. We take pride in sponsoring this special issue on obesity, which furthers the discussion on treating this critically important, complex disease.

Jonathan K.C. Knowles, Ph.D.
President of Global Pharma Research
The Roche Group



Finding that gene

Genetic mapping, western blotting and DNA probes feature.



West way: all you need for western blotting.

Western blot reagents

NEN Life Science www.nen.lifesci.com

A stable of products for western blots

NEN aim to offer a "one-stop shop" for Western blotting reagents. The product line includes Renaissance Western Blot Chemiluminescence Reagent Plus for horseradish peroxidase-catalysed chemiluminescence detection. The *Plus* version, developed from the original Renaissance reagent, provides increased sensitivity, and good signal-to-noise performance on both PVDF and nitrocellulose membranes. One kit supplies reagent covering up to 1,000 cm² of membrane. Reagents have a shelf-life of up to 12 months and can be stored at room temperature for at least eight weeks.

Reader Service No. 100

Mutation detection

Roche Diagnostics www.roche.com

Straight to the point with PCR

A new kit for the detection of the Factor V Leiden point mutation has been developed for use with the Lightcycler. In a combined procedure, amplification by PCR of a specific fragment of the factor V gene is followed by finer detection using a pair of hybridization probes, one of which is labelled to produce a fluorescent signal to indicate the presence of the point mutation. The kit includes primer and hybridization probes, pre-mixed reagents and a control template. Template DNA is simply added to each reaction mix, performed in a 20-μl volume.

Reader Service No. 101

HybQUEST reagents

PanVera Corporation www.panvera.com

Non-radioactive DNA labelling

Sensitive, non-radioactive detection of specific DNA-hybridization events is the claim made

for PanVera's HybQUEST Complete System. The system is suitable for membrane-based applications, including Southern blots, bacterial colony and plaque hybridizations, and dot or slot blots. It includes all reagents needed for labelling, hybridization and chemiluminescent detection of DNP-labelled DNA probes. The Label IT reagent enables one-step, non-destructive covalent labelling of DNA. The labelling procedure also allows the density of labels on the probe to be modified.

Reader Service No. 102

SeeBlue Plus 2

Novex

www.invitrogen.com

The name says it all

This new pre-stained marker from Novex has two non-blue-coloured proteins for unambiguous band identification. Pre-stained proteins in the molecular-weight range 3K to 188K resolve as sharp, clear bands on polyacrylamide gels. The marker comes ready to load and is recommended for monitoring western transfers.

Reader Service No. 103

Aquarius liquid DNA probes

Cytocell

www.cytocell.co.uk

Flexible FISH testing

A new range of liquid DNA probes offers improved multiple-sample screening for chromosomal abnormalities. Supplied in kit form, Aquarius contains specified probe, counterstain and a step-by-step users' guide. Kits for prenatal analysis with AneuCyte prenatal probes are now available for testing samples for sex-chromosome abnormalities and trisomies 21, 18 and 13. Kits are supplied in either 10- or 20-test formats.

Reader Service No. 104

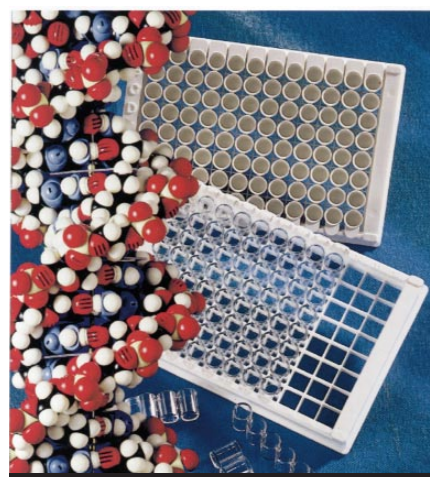
Reactive Polystyrene Surface

Corning Costar

www.corningcostar.com

Improved microplate hybridization

Use of nucleic-acid hybridization in a microplate has been limited by the difficulty in achieving effective immobilization of DNA onto the microplate surface. Corning's new Reactive Polystyrene Surface (RPS) is a high-binding surface that can be used as a solid substrate for binding amine-modified DNA oligonucleotides and subsequent hybridization. It has a high signal-to-noise ratio and offers rapid and specific binding of the capture probe, excellent binding and hybridization efficiency, stable bonds and well-to-well reproducibility. Suggested applications include colorimetric, luminescence



In the clear: RPS is also available in black or white.

and fluorescence. RPS is available as 1 × 8 stripwell plates in clear, white or black formats.

Reader Service No. 105

Vitotox

LabSystems Affinity Sensors

www.labsystems.fi

Rapid detection of DNA damage and gene mutations

The new Vitotox test uses bioluminescence to detect genotoxicity and cytotoxicity in just four hours. *Salmonella typhimurium* that has been genetically modified to carry the *Vibrio fischeri* luciferase operon emits light in response to genotoxicity, toxicity and mutagenicity of the sample. The test can be used in 96- or 384-well microtitre plates. Suggested applications include pre-screening to speed up the product-discovery phase in laboratories currently using the Ames test and validation of environmental compliance.

Reader Service No. 106

ADVERTISEMENT

TRANSFECTION REAGENTS

NEW TRANSOME
SUV Liposome Reagent

GENETRANSFER
MLV Liposome Reagent

HMG-1,-2 MIXTURE
Efficient Gene Carrier

Wako

Wako USA
877-714-1921

Wako GmbH
(49)02131 3110

www.wako-chem.co.jp

READER ENQUIRY NO. 37



Map maker: the CEQ2000 automation system.

CEQ 2000

Beckman Coulter

www.beckman.com

Automated DNA genotyping

Fragment analysis using the CEQ 2000 DNA Analysis System now includes automated DNA genotyping using microsatellite analysis. DNA fragments are separated by capillary electrophoresis using eight capillaries concurrently. Researchers can multiplex many samples simultaneously in each of the capillaries. An internal size standard ensures reproducible sizing. Accompanying software calls alleles based on fragment sizes. The system is recommended for applications involving genetic linkage analysis and genetic mapping.

Reader Service No. 107

GeneLine

Spectronics

www.spectroline.com

Gel documentation systems

The GeneLine range includes GeneLink and GeneCam, both with a CCD camera with 23 × 28 cm hood, and image acquisition software. By placing the hooded camera on the transilluminator over the gel, the image appears instantly on a monitor, and can be printed or saved. GeneCam includes GeneTools software to automatically locate lanes, bands, and edges and produce track profiles and results that can be exported into Excel.

Reader Service No. 108

Restore

Pierce

www.piercenet.com

Odour-free assays?

Restore western blot stripping buffer eliminates the need to rerun electrophoresis gels to optimize assay conditions for one western blot. It is supplied ready to use, is mercaptan-free and removes all immunoassay reagents in five to 15 min at room temperature. Membranes can be stripped at least four times without target protein loss.

Reader Service No. 109



Look under the hood: GeneLink, complete.

FirstChoice RLM-RACE Kit

Ambion

www.ambion.com

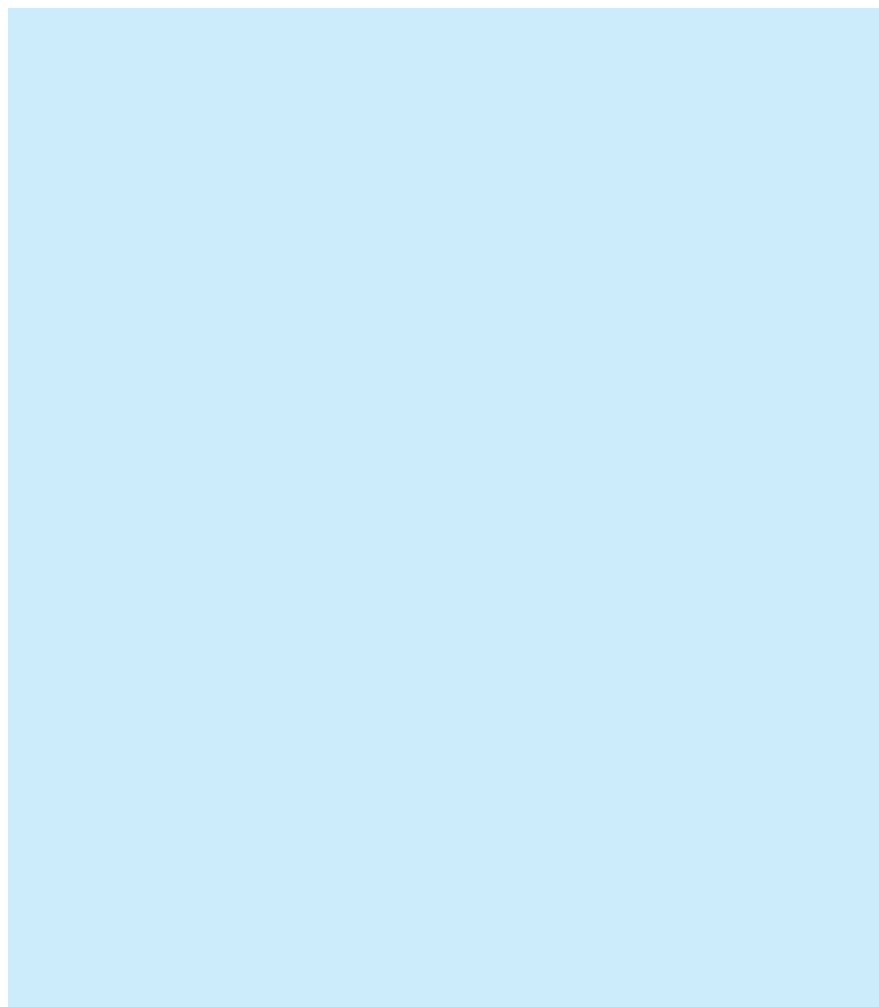
5'-end messenger RNA isolation

By preferentially selecting for capped mRNAs, this kit allows rapid amplification of the 5'-end of rare messages, even on a background of partially degraded mRNA, rRNA or genomic DNA. The kit comes with enzymes and reagents for five RLM-RACE cDNA preparations and primers for 100 PCR reactions.

Reader Service No. 110

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS



Check out
our homepage at
<http://www.resgen.com>

**BACs
YACs
PACs
cDNAs**

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 6

Look no further
for the best value in
Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!!

SIGMA
GENOSYS
(+44) (0) 1223 839000 • Fax (+44) (0) 1223 839200
email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 33

BYTING THE BULLET

Computers will play a major role in the biology labs of tomorrow

p683

IN SEQUENCE

Proteomics is revolutionizing protein discovery

p684

SKILLS SINK

US universities do the training — but industry is draining the talent

p686

DATA DILEMMAS

More specialists are needed to make sense of sequencing information

p687

Next-generation biologists must straddle computation and biology

Biological sciences are being transformed by the huge amounts of data emerging from newly sequenced genomes. But do the skills exist to cope?

Like a circus performer juggling daggers, chairs and flaming clubs, the bioinformatics field stands amid elements ranging from human genetics and clinical medicine, to biophysical studies of three-dimensional structure and studies of fruit-flies, yeast and bacteria.

As the field evolves and expands, so too do the skill sets necessary to excel in it. To attain competence demands new ways of thinking about knowledge and education, both by students and universities.

"The basic mentality clash between data-driven scientists and model-driven computer and math people stems from world views so different that all too often they can't have a conversation that makes sense," says Mark Perlin, chief executive officer of the genomics company Cybergenetics.

Different world views must coexist in one person to bridge the gulf that separates computer scientists from life scientists, says Perlin. Because phenomena and data are indivisibly associated, he says, there's no point in having people analyse terabytes of data if they have never spent time in the laboratory seeing what the data actually are and how they are obtained. "Until you've held a pipette in your hand, or run a PCR or a gel, you can't understand what the failure modes are going to be," Perlin explains.

Embrace the unknown

How can one achieve multidisciplinary competence in a rapidly changing field? Perlin asserts: "Those who seriously want to get into this field need to embrace what they don't know. Take courses in things you never studied before while working on some research project, and after some time you'll begin to understand what the people around you are talking about."

Quality interdisciplinary training doesn't come easily, though. It takes a month to introduce people to problems and two years to train a technician to do useful programming or lab work, but longer still to train a principal investigator.



Multiple-degree dynamics

When he considered augmenting his internal medicine studies with a mathematics PhD in 1978, Mark Perlin, now chief executive officer of Cybergenetics, a Pittsburgh-based genomics software firm, didn't exactly receive encouraging words. "Computer science and mathematics are not, and never will be, relevant to medicine," the dean of the University of Chicago told him then.

How times change. The path that Perlin chose over 20 years ago could be a template for today's bioinformatics professional. Perlin, who is also an adjunct faculty member at Carnegie-Mellon University and the University of Pittsburgh, earned a PhD in mathematics from the City University of New York, an MD from the University of Chicago, and a PhD in Computer Science from Carnegie-Mellon University.

His decision to pursue mathematics and computer programming along with medicine seemed logical, not trail-blazing. "Math broadened my view of medicine, so it seemed natural to interrupt medical studies midway through and study math for three years," he says.

He has no regrets now, but notes that some of that academic inflexibility that he faced initially still exists. "Biology is changing rapidly, but university departments are persistent forms, and rewards are distributed along department lines," he says. Those boundaries stifle innovation: "I have more freedom for genomics research in a start-up company than in a university department."

But earning that freedom didn't come easily. After medical school, he did a six-month postdoctoral stint at IBM's T. J. Watson Research Center in Yorktown Heights, New York, a one-year transitional medical residency in Pittsburgh, then joined the Carnegie-Mellon University faculty in computer science. While there, he earned a PhD in computer science, in the hope of embarking on an academic career.

But when he realized that academia could be "limiting", he launched his own company. "I started Cybergenetics in 1994 because I had a sense that the private small business model might be a better way to innovate and do research."

P.W.

Traditional departmental lines must become more blurred, says Jehoshua Bruck, a professor of computer science and electrical engineering at Caltech, in Pasadena, California. For example, the ideal next-generation biologist should develop both wet-lab expertise and software-writing ability.

Teamwork is key

Reorganizations along these lines are already under way at some institutions. The new Clark Center at Stanford University integrates traditionally separate disciplines, such as engineering, molecular biology, physics and computer science, to work on problems in common. Politically and financially the long-term viability of research

work groups, as opposed to more familiar departments, is likely to depend on adjusting how tenure is awarded and how funding and authorship priorities are allocated.

This more interdisciplinary, data-driven approach to biology has already begun producing results. Before Christoph Sensen, manager of the Canadian Bioinformatics Resource, in Halifax, Nova Scotia, saw a draft of the genome of *Sulfolobus solfataricus*, he thought that, on the basis of ribosomal sequences, the organism's genome was well organized, with no space between the genes.

When he examined the *Sulfolobus* genome, he found large spaces between genes, as well as repeats comprising up to a third of the genome. "We learned that the

▶ dogma that the prokaryotes are made with total economy is not so true," Sensen says. "This has made us change the way we think and the way we operate in the lab."

Challenges ahead

Nevertheless, the field must develop further to be more useful. In important ways, genomics has so far been built on scaled-up versions of classical methods. Although satisfactory for the linear task of factory-style sequencing, these will not accommodate the more complex problems that are emerging from a burgeoning quantity of data from a variety of sources.

At present, computational approaches to deal with these new forms of data are developed in an *ad hoc* way, says Chris Lee, at the Bioinformatics Institute in the University of California, Los Angeles. For example, to cluster genes by expression patterns, computational biologists generally try to define a metric that says "these genes express pretty similarly", Lee says. But that approach lacks both a model of the underlying phenomena and a rigorous computational method that can consider all possible interpretations.

More mundane but more immediate obstacles also clutter the way ahead. George Poste, former chief science and technology officer at pharmaceuticals giant SmithKline Beecham, frets about the lack of consistent nomenclature, and of standards to integrate research, trials and clinical databases. "It's no good realizing ten years from now that the research data can't be migrated downstream," he says.

Francis Ouellette, director of the Bioinformatics Core Facility at the Center for Molecular Medicine, in Vancouver, British Columbia, points to the problem of working digitally on tables and figures in thousands of papers that today exist only in print.

Mistrust by the public also provides a potential worry. Reaction to privacy and ownership issues, now mostly latent, could erupt, with unfortunate consequences for research funding and the licensing of new technologies. For proof of this, one need only recall the setbacks that agricultural biotechnology has suffered in Europe by its failure to allay the public's concerns about genetically modified foods.

Despite the growth of bioinformatics, biology in the age of genomics still has a great distance to go.

Getting connected to bioinformatics

For more information on bioinformatics techniques and general resources see:

- Bioinformatics courses
linkage.rockefeller.edu/wli/bioinfocourse/
- Canadian Bioinformatics Resource
www.cbr.nrc.ca/
- Cambridge Healthtech meetings
www.healthtech.com/
- Cold Spring Harbor short courses
nucleus.cshl.org/meetings/2000c-info.htm
- GOLD: Genomes Online Database
geta.life.uiuc.edu/~nikos/genomes.html

- Human Genome Organisation
www.gene.ucl.ac.uk/hugo/
- IBM Computational Biology Center
www.research.ibm.com/topics/serious/bio/
- Kyoto Encyclopedia of Genes & Genomes
kegg.genome.ad.jp/kegg/
- National Center for Biotechnology Information
www.ncbi.nlm.nih.gov/
- National Center for Genome Resources
www.ncgr.org/
- UCLA Bioinformatics Institute
www.bioinformatics.ucla.edu/

Political conflict presents another obstacle to advancing the professions. The apparent unravelling of the public-private collaboration in genomics in the United States is demonstrated by recent letters between Francis Collins, director of the National Human Genome Research Institute, and Craig Venter, president of Celera Genomics (see www.bioinform.com).

Each side accused the other of bad faith in holding up their side of the bargain (see *Nature* **404**, 117; 2000). This exchange hints at the difficulties that lie ahead in reconciling public and private interests in a rapidly growing enterprise whose private arm is

already worth in the order of \$45 billion.

Despite the growth of bioinformatics, biology in the age of genomics still has a great distance to go. Christoph Sensen observes that genomic sequences dating from 1986 have open reading frames that are still of unknown function, and the number of these early sequences is dwarfed by those now emerging.

Resolving those problems will require a new breed of scientists who are comfortable both with the old lab skills and with new computational techniques. "To grasp the complexity will require much more wet-lab activity and much more analysis," Sensen concludes.

Potter Wickware

Companies of all sizes are prospecting for proteins

It is being called a land grab at the UK-based company Oxford GlycoSciences, but the property being prospected for is intellectual — patents on proteins. Fuelled by \$50 million raised at the end of February on the London Stock Exchange, the company is racing to patent as many proteins as possible during the next two years. In mid-March it announced that it had filed patent applications covering more than 800 different protein-use combinations.

Certainly, investors seem impressed by the company's aim, and its share price shot up from a little over £5 in early January to £32 in the second week of March, despite a warning from Michael Wandra, the chief executive officer, that extra investment is likely to be postponed to beyond 2002.

The multiplier in this equation is, of course, the Human Genome Project. The race to complete the human genome is nearing the end, and as the databases fill up with DNA sequences, and increasing numbers of genes are identified with greater certainty within those sequences, companies such as Oxford GlycoSciences and its rivals Large Scale Biology Corp. in the United States have

access to a powerful new tool for identifying proteins and protein complexes.

Why all the excitement and financial faith? Proteins orchestrate the life and death of all living organisms. They are the main product coded for by genes. When something goes wrong with a protein, it can cause disease. So identifying the proteins and protein complexes associated with all processes in healthy and diseased organisms is critical to an understanding of human biology. It is a gargantuan task, which only a decade or so ago was a highly specialized cottage industry. Now the task has acquired a new name — proteomics — and it is on the threshold of becoming a highly mechanized industry (see *Nature* **403**, 815–816; 2000).

Out of sequencing

During the cottage-industry days, a protein could be identified only by isolating it and then laboriously determining the sequence of its amino acids. Finding all the proteins with which it interacted and identifying them was several orders of magnitude more difficult. Now sequencing is not needed.

Instead, highly accurate mass spectrom-



Profit potential: proteomics could open the door to major commercial gains.

eters determine the mass of the protein or subunits of the proteins in a complex. If the proportion of the protein's constituent amino acids is known, computers can then take over the task of identifying the protein. Software then looks for a DNA sequence that would code for a protein made up of the known amino-acid composition. Finally, the mass spectrometer determines the protein's exact mass.

Compared with their peers of two decades ago, researchers applying these techniques to drug discovery are off to a flying start in the effort to identify protein function. Even if the databases they search are not annotated with function, the identified proteins are those taken from diseased tissue. So there is at least a correlation pointing to the protein's potential application.

There are still many weaknesses in the technologies used for proteomics (see *Nature* **402**, 715–720; 1999), and the subsequent job of identifying the exact role of the proteins in the cell remains a complex biological problem. But current technology is far enough advanced to meet the aims of Oxford GlycoSciences, according to Burns. The company can analyse 1,000 proteins per month and is spending the money raised in February on further developing its protein identification effort.

"We have optimized the first generation of the technology," says Burns. "It is cumbersome — like a Model-T Ford — but it gets us from A to B." That destination is the world's patent offices. Oxford GlycoSciences' strategy is driven by the perception that there is a single window of opportunity — now — when the DNA sequences and gene maps currently being completed can be exploited to identify proteins. "It is," says Burns, "two minutes to twelve and counting."

The aim of this frenetic activity is that in future years these patents will, so to speak, pay dividends. As knowledge of the proteins

and protein complexes in the body involved in disease deepens, those owning protein and/or gene patents could find themselves in a powerful commercial position.

Amino activity

Broadly speaking, proteomics can be divided into two conceptual approaches, argues Walter Blackstock, from Glaxo Wellcome Research and Development. These are expression proteomics — comparing the complement of proteins expressed by healthy and diseased tissue — and cell-map proteomics — the study of protein–protein interactions. Both routes rely on a suite of technologies that separate and identify proteins associated with disease, and both have the aim of improving the selection of the optimum target for drug development.

Industry has led academia in the field of proteomics, both in Europe and the United

States, according to Alf Game, head of the genetics and biochemistry branch of Britain's Biotechnology and Biological Sciences Research Council (BBSRC). Nevertheless, large pharmaceutical companies were not immediately impressed by the technology. Initially, spin-offs from universities led the way.

"There was a lot of hype," says Blackstock. "You have to remember that proteomics is not the only approach to drug discovery and target validation. We were weighing a non-proven technology against proven methods."

But the caution is fading. "Big pharma is now switching to proteomics," says Blackstock, who heads the cell-map proteomics unit at Glaxo Wellcome. The unit was formed 14 months ago after 12 months spent studying the field informally.

Proteomics is such a large and complex field that there is still room for newcomers to the industrial scene. It is the proteins themselves that make the field so challenging and unpredictable. Although the human genome codes for about 100,000 genes, the variety of proteins in the body is many times greater. This is because genes essentially code for an amino-acid sequence.

Once expressed, many things happen to these proteins. For example, functional groups are added or the protein is cleaved. Proteomics aims to identify proteins *in vivo*, where such post-translational changes might mean that there are millions to be detected. Furthermore, the state of proteins changes with time. Unlike genomics, which in essence had one genome per organism on which to focus, there is, says Burns, a nearly infinite number of proteomes.

This makes finding improved proteomics tools the main goal of both industrial and academic researchers. Developing



The power of prediction

Andrea Califanos, programme director of the Computational Biology Center and manager of the Functional and Structural Genomics Group at IBM's T. J. Watson Research Center in Yorktown Heights, New York, hopes that computation can make biology more predictable.

"Biology has been a descriptive science for a long time, and has lacked the abstract theoretical framework that allows you to draw complex inferences about phenomena, but this is changing," he says. "I think the union of the theoretical sciences — mathematics, physics, computer science — with biology is going to let us model, predict and, most importantly, understand the most intimate mechanisms of life."

This union — and the prediction that it will produce — may change the way in which experiments are designed. Califanos points to gene-expression analysis, where the synergy between statistics, mathematics and biology is leading to successes in relating genotype to phenotype.

Expanding this sort of success must entail more than developing a new generation of scientists with mixed degrees, he says, adding that "We need a new language". Organizations and centres such as the Center for Physics and Biology at Rockefeller University, the Institute for Advanced Studies at Princeton, and the Biophysics Institute at Johns Hopkins University, may provide one.

At IBM, rather than attracting mixed-degree people, Califanos has focused on building a team made up of an "extremely heterogeneous and eclectic group" of scientists.

P.W.

▶ these can provide the basis for a start-up company.

Upstart start-up

Take CellZome, which has spun off from the European Molecular Biology Laboratory (EMBL) in Heidelberg. This company is so new that it hasn't even got a website. Gitte Neubauer is a mass spectrometrists and one of six founding scientists from EMBL. Three are biologists, two are mass spectrometrists and one is a bioinformatician.

Using novel technology developed by Bertrand Seraphin (one of the co-founders of CellZome), the company plans to purify complexes of interacting proteins and identify them using mass spectroscopy and bio-informatics. It will complete the work by carrying out the biochemistry to validate its conclusions. The company is currently looking for collaborators interested in applying its techniques and scientific know-how to drug discovery.

Taken together, Oxford GlycoSciences, Glaxo Wellcome and CellZome give a good sense of the vibrancy of the newly emerging field of proteomics. Although they vary in both size and age, all of them are acutely aware of the need for skills in mass spectroscopy and bioinformatics. People armed with those skills should find themselves in demand.

Helen Gavaghan

Further information

For a good sense of the techniques used in proteomics see:

1. Kell, D. B. & King, R. D. *Trends Biotechnol.* **18**, 93–98 (2000).
2. Brent, R. *Cell* **100**, 169–183 (2000).
3. Abbott, A. *Nature* **402**, 715–720 (1999).
4. Blackstock, W. P. & Weir, M. P. *Trends Biotechnol.* **17**, 121–127 (1999).

For proteomics opportunities and information see:

● Oxford GlycoSciences — jobs and technologies
<http://www.ogs.com>

● Glaxo Wellcome
<http://science.glaxowellcome.com/start.htm>

● SmithKline Beecham
<http://www.sb.com>

● Incyte Genomics offers surfers a proteomics tour
<http://www.incyte.com/science/>

● Human Proteome Initiative
<http://ebi.ac.uk>

● Proteomics, protein databases and bioinformatics
<http://www.expas.ch> and <http://www.isb-sib.ch>

● Links to academic proteomics sites
http://www.proteome.med.umich.edu/links/proteo_sites.html



Growth industry: understanding protein function is key to future drug discovery.

Training

United States gives priority to skills shortage

Bioinformatics marries together a wide range of scientific disciplines, but with a global shortage of skilled researchers, training is high on the agenda.

Washington

Industry is draining bioinformatics talent from universities faster than it can be replenished. This is good news for the people getting such training, but bad news for the institutions that are scrambling to provide it, says Francis Ouellette, at the University of British Columbia's Center for Molecular Medicine and Therapeutics. Ouellette and Christoph Sensen at Canadian Bioinformatics Resource, in Halifax, Nova Scotia, run a four-part survey series (one week each for bioinformatics, genomics, proteomics and tools development), which introduces people to the field. Ouellette worries that the series is only a temporary patch on the problem.

Sensen stresses the difficulties academic groups have in finding and retaining talent. "In two years of looking I haven't found a person willing to be an understudy in this environment. PhDs either go to a company or to a nice warm place in the United States where they also get more money. But there is an urgent need for more PhDs in academia because that's where much of the real science is done."

Chris Lee, of the Bioinformatics Institute at the University of California, Los Angeles, concurs. Industry has the data, he says. But it

lacks the "intellectual diversity" of a full-service university, as well as the freedom to "sit around talking about problems with people from different backgrounds".

The gap between supply and demand in bioinformatics is receiving official recognition in the United States. The US National Institutes of Health (NIH) supports bioinformatics mainly through two institutes, the National Human Genome Research Institute and the National Library of Medicine. However, for the field to grow, centres outside the NIH must also arise. The NIH approves the concept of developing such "centres of excellence", but has been slow to commit funding and develop infrastructure.

The National Institute of General Medical Sciences has also committed itself to

The gap between supply and demand in bioinformatics is receiving official US recognition.

funding training slots, and a fourth branch of the NIH, the National Center for Research Resources (which is not an institute), has put itself behind shared bio-computational resources at more than a dozen centres nationwide. The Department of Energy's Argonne and Oak Ridge laboratories are also huge funders of bioinformatics work, as is, to a somewhat smaller extent, the Department of Defense.

On the private side, the Howard Hughes Medical Institute (HHMI) has declared that

Borrowing methods and models



Kimmen Sjölander's early work in using Hidden Markov Models (HMMs) in genomics illustrates how bioinformatics reaches backwards to move forwards. HMMs originated in machine learning and were developed for speech

recognition, but they turned out to be "fantastically successful" for modelling proteins, says the scientist from Celera Genomics, Foster City, California. "This typifies what bioinformatics is all about, to recognize that a method in one context can be transferred to another."

People with these skills are valuable to industry. "Pharma companies are desperately trying to organize and prioritize their drug targets," Sjölander says. "To do this you begin by making inferences about functions that are embodied in sequence."

However, that skill does not come easily. To develop fully the ability to make inferences, one needs to feel comfortable moving back and forth between computational and experimental data. "It's important for people who want to work in this field to have a true basis of understanding in more than one field. In addition to molecular biology, math, statistics and some elements of machine learning would be helpful," she says. "You can't do it with knowledge of just one specialty."

Sjölander thinks that bioinformatics should be a science in its own right. The demand certainly exists. The sheer quantity of data being generated from such a wide variety of sources requires the constant creation of new methods using mathematical or statistical modelling. "There's so much data being generated and we need to have methods that will respond to its particularities."

To increase the supply of computational biologists to deal with the demand being created by the data deluge, Sjölander believes that a cultural change is in order. "Historically, biologists haven't felt comfortable talking to mathematicians and computer scientists, who in turn don't really understand a lot of the issues in biology," she says. "That has to change."

P.W.

it will appoint investigators in computational biology, a departure for an organization that until now has avoided funding research in what it viewed as engineering disciplines. Now, however, it is becoming clear that bio-computation is not only part of HHMI's biomedical mission, but is one of its most critical elements.

Other support is also issuing from the Alfred P. Sloan Foundation, which has recently called for proposals to fund academic units that create professional master's degrees in biology. Traditionally, these degrees have not carried the same weight in biology as in engineering or business, where they are terminal degrees with full professional credentials. A two-year training cycle might help to relieve industry's acute need, where competent support staff, rather than technically savvy principal investigators, is what is needed.

Recognition of need

The NIH and the HHMI, respectively the largest public and private funders of biomedical science in the United States, have both previously acknowledged the need for more training in computational biology. But the money to pay for large-scale programmes has not materialized. Now the two agencies have signalled that this may change.

Last June, an NIH working group recommended a sweeping national programme to promote computational biology, its cornerstone being funding for up to 20 national

centres (see <http://www.nih.gov/about/director/060399.htm>). Although the goals of this Biomedical Information Science and Technology Initiative (BISTI) have not been met, Carol Dahl, NIH assistant to the director of special technologies, remains optimistic.

BISTI will be considered by all NIH institute directors this spring, and money for the programme could be in the budget for 2001. President Bill Clinton's 2001 budget proposal contains more money for large centres rather than new money for additional investigator-initiated grants.

The HHMI has sent similar signals. When Tom Cech was appointed director of the institute, he was quick to single out computational biology as an area in which it might increase its support (see *Nature* **398**, 256; 1999). But, because of the necessary shift towards collaboration and data collection, academia may need to rethink the way such work is rewarded, Cech warns. Tenure might not come easily to a researcher whose credit to a paper is lost in a sea of names. Other systems of acknowledgement, where each person's individual contributions are noted, may become necessary as authorship lists to collaborative papers become longer.

Last winter, the HHMI hosted a symposium to examine how computation and biology could be better merged and funded, from both public and private sources. However, according to an institute spokesman, no new funds have been committed for computational biology.

Potter Wickware

Training

Europe seeks solution to bioinformatics shortfall

London

A shortage of skills in bioinformatics is not limited to the United States. Europe is also feeling the pinch. The need for bioinformaticians to make sense of the data emerging from the Human Genome Project is well known among the industrialists and academics working in the field. It is something that was emphasized repeatedly to Lesley Thompson, who runs the life sciences interface programme of the Engineering and Physical Science Research Council, as she toured the United Kingdom's universities to identify ways in which the physical sciences could help life sciences.

For proteomics, bioinformaticians are, of course, essential, but they are not very easy to find. Both Walter Blackstock, of Glaxo Wellcome, and Gitte Neubauer, of the start-up company CellZome in Heidelberg, say

that when recruiting for their proteomics effort they received only ten applicants for posts in bioinformatics, compared with 100 for biology.

One way to begin to tackle this imbalance, says Graham Cameron, joint head of the European Bioinformatics Institute (EBI) in Cambridge, an outpost of the European Molecular Biology Laboratory, would be to make sure that training in bioinformatics is included in all undergraduate biology courses. "Bioinformatics needs to be so pervasive in biology that it transcends its geeky image," he says.

Another way of addressing the imbalance is by establishing more academic programmes at all levels. Five of the United Kingdom's research councils have banded together to strengthen research in bioinformatics. The Biotechnology and Biological

Mining databases will be crucial in the coming years as researchers try to pin down protein and gene function.

Sciences Research Council (BBSRC), the Engineering and Physical Sciences Research Council (EPSRC), the Medical Research Council (MRC), the National Environment Research Council (NERC) and the Particle Physics and Astronomy Research Council (PPARC) will pool their resources so that three universities will receive £1.25 million (US\$2 million) each over five years. The money will pay for a professorship, technical support and some infrastructure.

Banking on information

The EBI itself is primarily an archiving service. It looks after the gene and protein data that are emerging worldwide from the studies of the genomes of all organisms. Some 20 per cent of the institute's activity is reserved for research methods to improve the presentation and utility of its unique set of databases. Both nucleotides and proteins fall under the institute's remit, with the protein work being a collaboration with the Swiss Institute of Bioinformatics (SIB) in Geneva.

This month the two institutes are releasing a new database that comprises all vertebrate sequences that have a known function. The database is part of the institutes' Human Proteomics Initiative. If someone is seeing a sequence of human protein — derived from a gene identified in the Human Genome Project — but has no idea what the protein does, the researcher will be able to compare the new sequence with the database of vertebrate sequences. If they find a similar sequence or sequences from other species, then the researcher will have been given a clue as to what the function of the new sequence might be in humans.

Data mining

The new database is made up of information already in the public domain — in databases maintained by EBI and SIB — but it is reorganized into what is known as a cluster database. This kind of database structure allows information with differing degrees of commonality to be grouped in clusters. It is not a new tool, but it is this sort of tool that bioinformaticians develop in their efforts to find potentially meaningful correlations: hence the need for both biological and computational knowledge.



Cameron: to succeed, bioinformatics must transcend its geeky image.

The most important protein database that the EBI and SIB maintain — its information was mined for the new vertebrate database — is SWISS-PROT. This has been growing since 1986 and includes sequence data from all organisms, derived primarily from gene sequences.

Curators then ensure that there are no duplicate sequences, and research the literature to find out as much as they can about the function of the sequences. This additional information is attached as annotations to the relevant sequences. The process is time-consuming, but it makes SWISS-PROT a valuable research tool. It is free to academics, but industry pays a licence fee — the money pays for additional curators and researchers.

Aware of the need to support efforts to turn the data emerging from the Human Genome Project into applications, the EBI and SIB have stepped up efforts to annotate the human protein sequences. They are also

looking at ways to provide researchers with a database of protein-protein interactions, something that researchers such as Mathias Uhlen, head of the department of biotechnology and biochemistry at the Royal Institute of Technology in Sweden, believes is essential.

Clearly, mining databases will be crucial in the coming years as researchers try to pin down protein and gene function or spin off companies from their university departments, and as industry looks for scientific insight and checks that there is no prior art to invalidate any patent applications. Says Cameron: "The winners in the next year or so will be those who do the bioinformatics right."

Helen Gavaghan

Further information

- The European Bioinformatics Institute:
<http://www.ebi.ac.uk>
- The Swiss Bioinformatics Institute
<http://www.isb-sib.ch>
- The EPSRC
<http://www.epsrc.ac.uk>
- The BBSRC
<http://www.bbsrc.ac.uk>
- The NERC
<http://www.nerc.ac.uk>
- The MRC
<http://www.mrc.ac.uk>
- The PPARC
<http://www.pparc.ac.uk>



Sequence searching: databases are powerful tools for understanding proteins.